PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: KAI KROHN, ET AL.

Serial No.: 09/508,658 Group No.: 1634

Filed: NOVEMBER 3, 2000 Examiner: J.S. SITTON

For: NOVEL GENE DEFECTIVE IN APECED AND ITS USE

Attorney Docket No.: U 012653-9

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

REQUEST FOR DECLARATION OF INTERFERENCE PURSUANT TO 37 CFR 41.202

In accordance with the provisions of 37 CFR 41.202 applicants suggest that an interference be declared:

CERTIFICATE OF MAILING/TRANSMISSION (37 CFR 1.8a)

I hereby certify that this correspondence is, on the date shown below, being:

MAILING

☒ deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to the Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450

FACSIMILE

☐ transmitted by facsimile to the Patent and Trademark Office to **(571)-273-8300**

Janet Cord
Signature

Date: September 29, 2006

Janet I. Cord
(type or print name of person certifying)

(1) between the subject application 09/508,658 and U.S. Patent 6,951,928 (Peltonen - Issue date of Patent- October 4, 2005). This provides sufficient information to identify the patent with which the applicant seeks an interference. (37 CFR 41.202(a)(1)).

A request for declaration of interference pursuant to 37 CFR 1.41.202 between the divisional of this application, US patent application 11/501,979 and U.S. patent 6,951,928 is being filed in US patent application 11/501,979. Both of these requests are being filed within one year of the issue date of US patent 6,951,928.

(2) The applicants believe claims 27 and 36 presently on file in the subject application interfere with claims 2, 3, 4 and 8 of U.S. Patent 6,951,928.

Applicants propose the following counts:

Count 1:

An isolated nucleic acid molecule comprising nucleotides 1-2020 of nucleotide sequence SEQ ID NO: 1 of U.S. Patent 6,951,928 or nucleotides 17-2036 of SEQ ID NO: 1 of U.S. Patent Application 09/508,658.

Count 2:

An isolated nucleic acid molecule differing from the nucleic acid sequence of SEQ ID NO: 1 of U.S. Patent 6,951,928 by a substitution, wherein the substitution is a change of cytosine to thymidine at nucleotide position 889 or an isolated nucleic acid molecule differing from the nucleic acid sequence of SEQ ID NO:1 of U.S. Patent Application 09/508,658 by a substitution, wherein the substitution is a change of cytosine to thymidine at nucleotide position 905.

The claims of the parties that correspond to Count 1 are :

U.S. Patent Application 09/508,658	27
U.S. Patent 6,951,928	2, 3 and 4

The claims of the parties that correspond to Count 2 are:

U.S. Patent Application 09/508,658	36
U.S. Patent 6,951,928	8

In this paragraph, applicants have identified all claims the applicants believe interfere, have proposed one or more counts, and have shown how the claims correspond to one or more counts (37 CFR 41.202(a)(2)). Additional information describing how claims correspond to one or more counts is found in the next section of this request for declaration of interference.

3) Claim Chart

For each count, a claim chart is provided comparing at least one claim of a party corresponding to the count.

A showing why the claims interfere within the meaning of §41.203(a) is provided. (37 CFR 41.202(a)(3)).

Count 1	US patent application 09/508,658	US patent 6,951,928	Why the claims interfere
	Claim 27	Claim 2	

An isolated nucleic acid molecule comprising nucleotides 1-2020 of nucleotide sequence SEQ ID NO: 1 of U.S. Patent 6,951,928 or nucleotides 17-2036 of SEQ ID NO: 1 of U.S. Patent Application 09/508,658.	An isolated nucleic acid molecule comprising the SEQ ID NO:1.	The isolated nucleic acid molecule of claim 1 wherein the molecule is DNA or RNA. Claim 3. An isolated nucleic acid molecule according to claim 1 wherein the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO:1. Claim 4. The isolated nucleic acid molecule of claim , wherein the nucleic acid molecule consists of the contiguous nucleotide sequence of SEQ ID NO:1, or the coding region thereof that encodes the polypeptide of SEQ ID NO:2.	SEQ ID NO:1 of Claim 27 of US patent application 09/508,658 includes nucleotides 1-2020 of SEQ ID NO:1 claimed in claims 2, 3 and 4 of U.S. Patent 6,951,928. SEQ ID NO:1 of Claim 3 of US patent 6,951,928 includes nucleotides 17-2036 of SEQ ID NO:1 of claim 27 of US patent application 09/508,658.

According to 37 CFR 10. 41.203(a) an interference exists if the subject matter of a claim of one party would, if prior art have anticipated or rendered obvious the subject matter of a claim of the opposing party and vice versa.

Claim 27 of the subject application interferes with claims 2, 3 and 4 of U.S. Patent 6,951,928 and vice versa, because each claims a nucleotide sequence of SEQ ID NO: 1. The nucleotide sequence of these SEQ ID Nos: 1 that encode the polypeptide of SEQ ID NO: 2 of the respective SEQ ID NOs: 1 is the same. The nucleotides 17-2036 of SEQ ID NO: 1 included in claim 27 of this application which is the sequence encoding the amino acid sequence of SEQ ID NO:2 is the same as nucleotides 1-2020 of SEQ ID NO: 1 of claim 3 of U.S. Patent 6,951,928 which is the sequence encoding the amino acid sequence of SEQ ID NO:2. SEQ ID NO:2 of US patent application 09/508,658 is the same as SEQ ID NO:2 of U.S. patent 6,951,928. For anticipation, there must be no difference between the claimed invention and the reference disclosure, as viewed by a person of ordinary skill in the field of the invention. *Scripps Clinic & Res. Found. v. Genentech, Inc.*, 927 F.2d 1565, 18 USPQ2d 1001 (Fed. Cir. 1991). Therefore, since the coding sequences of SEQ ID NOs:1 of the respective patent application and patent are the same, the subject matter of the claim 27 of US patent application 09/508,658 and claims 2, 3 and 4 of US patent 6,951,928, anticipate the coding sequence of each other and if not considered to anticipate, each other would certainly be obvious in view of the other claim(s)

Count 2	US patent application 09/508,658	US patent 6,951,928	
	Claim 36	Claim 8	

An isolated nucleic acid molecule differing from the nucleic acid sequence of SEQ ID NO: 1 of U.S. Patent 6,951,928 by a substitution, wherein the substitution is a change of cytosine to thymidine at nucleotide position 889 or an isolated nucleic acid molecule differing from the nucleic acid sequence of SEQ ID NO:1 of U.S. Patent Application 09/508,658 by a substitution, wherein the substitution is a change of cytosine to thymidine at nucleotide position 905.	An isolated nucleic acid molecule comprising SEQ ID NO:1 wherein the nucleotide at position 905 is a T instead of a C.	An isolated nucleic acid molecule differing from the nucleic acid sequence of SEQ ID NO:1 by a substitution, wherein the substitution is: changes of cytosine to thymidine at nucleotide position 889, guanosine to thymidine at nucleotide position 358, adenosine to guanosine at nucleotide position 374, guanosine to adenosine at nucleotide position 1052, or cytosine to adenosine at nucleotide position 1094.	Both claim 36 of US patent application 09/508,658 and claim 8 of US patent 6,951,928 define the same mutation in the same location of the nucleotide sequence encoding the polypeptide of SEQ ID NO:2. This is shown in the attachment 1.

Claim 36 of US patent application 09/508,658 interferes with claim 8 of U.S. Patent 6,951,928 and vice versa because they define the same mutation. This is the mutation at nucleotide 768 of the coding sequence. As stated above, for anticipation, there must be no difference between the claimed invention and the reference disclosure, as viewed by a person

of ordinary skill in the field of the invention. *Scripps Clinic & Res. Found. v. Genentech, Inc.*, 927 F.2d 1565, 18 USPQ2d 1001 (Fed. Cir. 1991). Therefore, the subject matter of these claims anticipate each other and if not considered to anticipate each other would certainly be obvious in view of the other claim.

4) The applicants will prevail on priority because U.S. application 09/508,658 is a 35 USC 371 application of PCT application FI98/00749 filed on September 23, 1998 which claims priority from Finnish patent application 973762 filed on September 23, 1997.

Finnish patent application 973762, the priority application, describes SEQ ID NO: 1 the subject matter claimed in claim 27 on pages 19-22 of the application.

Support for claim 36 is found, inter alia, on page 5, lines 28-32; page 8, lines 35-36 and Fig. 3(b) of the Finnish priority application.

A copy of Finnish application 973762 is attached for the Examiner's convenience.

U.S. Patent 6,951,928 is a 35 USC 371 application of PCT EP98/06294 filed on October 2, 1998. Priority is claimed from German patent applications DE 97 11 7154; DE 97 11 7398; and DE 97 11 9810 filed on October 2, October 8 and November 12, 1997 respectively, all of which were filed after September 23, 1997, the date of applicants' Finnish patent application 973762.

Applicants earliest constructive reduction to practice of September 23, 1997 is earlier than the earliest priority date of U.S. Patent 6,951,928.

Therefore, as the applicants have the earliest filed application that includes a description of the interfering subject matter and their Finnish patent application 973762, has the earliest filing date which is evidence of the earliest constructive reduction to practice,

applicants will prevail on priority.

(5) No claims have been added or amended to provoke an interference. Claim 36 was included in the last amendment filed in US patent application 09/508,658.

(6) The following is a chart showing where the disclosure provides a constructive reduction to practice within the scope of the interfering subject matter (37 CFR 41.202(a)(6)).

		US patent application 09/508,658	Finnish patent application 973762
An isolated nucleic acid molecule comprising nucleotides 1-2020 of nucleotide sequence SEQ ID NO: 1 of U.S. Patent 6,951,928 or nucleotides 17-2036 of SEQ ID NO: 1 of U.S. Patent Application 09/508,658.	Claim 27 of US patent application 09/508,658	See sequence listing of SEQ ID NO:1 on pages 21-24 of the application.	See sequence listing on pages 19-22 of the application

An isolated nucleic acid molecule differing from the nucleic acid sequence of SEQ ID NO: 1 of U.S. Patent 6,951,928 by a substitution, wherein the substitution is a change of cytosine to thymidine at nucleotide position 889 or an isolated nucleic acid molecule differing from the nucleic acid sequence of SEQ ID NO:1 of U.S. Patent Application 09/508,658 by a substitution, wherein the substitution is a change of cytosine to thymidine at nucleotide position 905.	Claim 36 of US patent application 09/508,658.	Support for claim 36 is found, inter alia, on page 5, lines 16-20, page 8, lines 22-23 and Figure 3(b) of the US patent application.	Support for claim 36 is found, inter alia, on page 5, lines 28-32; page 8, lines 35-36 and Fig. 3(b) of the Finnish priority application.

It is respectfully requested that the interference be declared.

Respectfully submitted,

JANET I. CORD
LADAS & PARRY
26 WEST 61ST STREET
NEW YORK, NEW YORK 10023
REG. NO.33778 (212)708-1935

PATENTTI- JA REKISTERIHALLITUS
NATIONAL BOARD OF PATENTS AND REGISTRATION

Helsinki 20.10.98

E T U O I K E U S T O D I S T U S
P R I O R I T Y   D O C U M E N T

Hakija Applicant	FINNISH IMMUNOTECHNOLOGY LTD Tampere
Patenttihakemus nro Patent application no	973762
Tekemispäivä Filing date	23.09.97
Kansainvälinen luokka International class	C 12N
Keksinnön nimitys Title of invention	

"Novel gene"
(Uusi geeni)

Täten todistetaan, että oheiset asiakirjat ovat tarkkoja jäljennöksiä patentti- ja rekisterihallitukselle alkuaan annetuista selityksestä, patenttivaatimuksista, tiivistelmästä ja piirustuksista.

This is to certify that the annexed documents are true copies of the description, claims, abstract and drawings originally filed with the Finnish Patent Office.

Satu Vasenius
Jaostopäällikkö

Maksu 385,- mk
Fee 385,- FIM

Osoite: Arkadiankatu 6 A
Address: P.O.Box 1160
Puhelin: 09 6939 500
Telephone: + 358 9 6939 500
Telefax: 09 6939 5204
Telefax: + 358 9 6939 5204

1 2

Novel gene

Field of the invention

The present invention relates to a novel gene, a novel protein encoded by said gene, a mutated form of the gene and to diagnostic and therapeutic uses of the gene or a mutated form thereof. More specifically, the present invention relates to a novel gene defective in autoimmune polyendocrinopathy syndrome type I (APS I), also called autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) (MIM No. 240,300).

Background

Autoimmune polyglandular syndrome type I (APS I), also known as autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), is a rare recessively inherited disease (MIM No. 240,300) that is more prevalent among certain isolated populations, such as Finnish, Sardinian and Iranian Jewish populations. The incidence of the disease among the Finns and the Iranian Jews is estimated to be 1:25000 and 1:9000, respectively, whereas only few cases in other parts of the world are found each year.

APECED is one of the two major autoimmune polyendocrinopathy syndromes. The causing factor of APECED has not yet been identified. In APECED, the patient develops chronic mucocutaneous candidiasis soon after birth, and later several organ-specific autoimmune diseases, mainly hypoparathyreoidism, Addison's disease, chronic atrophic gastritis with or without pernicious anemia, and in puberty gonadal dysfunction occur [Ahonen P, Clin. Genet. 27 (1985) 535-542]. An accepted criterion for diagnosis of APECED is the presence of at least two of the three main symptoms, Addison's disease, hypoparathyroidism and candidiasis, in patients [Neufeld, M. *et al.*, Medicine 60 (1981) 355-362]. Immunologically, the major findings are the presence of high-titer serum autoantibodies against the effected organs, antibodies against *Candida albicans*, and

low or lacking T-cell responses toward candidal antigens [Blizzard, R. M. and Kyle M., J. Clin. Invest. 42 (1963) 1653-1660; Arulanantham, K. *et al.*, New Eng. J. Med. 300 (1979) 164-168; Krohn, K. *et al.*, Lancet 339 (1992) 770-773; Uibo R. *et al.*, J. Clin. Endocrinol. Metab. 78 (1994) 323-328]. The disease usually occurs in childhood, but new tissue specific symptoms may appear throughout life [Ahonen, P. et al., New Engl. J. Med. 322 (1990) 1829-1836]. APECED is not associated with a particular HLA haplotype, and both males and females are equally affected consistant with the autosomal recessive mode of inheritance.

The locus for the APECED gene has been mapped to chromosome 21q22.3 between gene markers D21S49 and D21S171 based on linkage analysis of Finnish families [Aaltonen, J. *et al.*, Nature Genet. 8 (1994) 83-87]. Recently, Börses *et al.* reported a maximum LOD score of 10.23 with marker D21S1912 just proximal to the gene PFKL, and thus by linkage disequilibrium studies the critical region for APECED can be considered to be less than 500 kb between markers D21S1912 and D21S171. Locus heterogeneity was not revealed by linkage analysis of non-Finnish families [Björses, P. et al., Am. J. Hum. Genet. 59 (1996) 879-886].

Physical maps of human chromosome 21q22.3 have been developed using YACs, and bacterial based large insert cloning vectors [Chumakov et al., Nature 359 (1992) 380; Stone et al., Genome Res. 6 (1996) 218], and many laboratories have contributed to the construction of a transcription map of the whole chromosome and 21q22.3 in particular [Chen *et al.*, Genome Res. 6 (1996) 747-760; Yaspo *et al.*, Hum. Mol. Genet. 4 (1995) 1291-1304]. Numerous trapped exons from chromosome 21 specific cosmids and also physical contigs from the APECED critical region have been identified and partially characterized. In addition, a number of ESTs from the international human

genome project have been mapped to the APECED critical region.

Recently, as part of the international efforts of generating the entire sequence of human chromosome 21 and international agreements on the immediate availability of this type of sequence data, the partial sequence of the APECED gene critical region was made available in GenBank by the Stanford Human Genome Center which is currently carrying out the sequencing of 1.0 Mb around the critical region of the APECED gene.

However, the precise location and the sequence of the APECED gene and the nature of the gene product have not so far been clarified. Thus at present the diagnosis of APECED is based mainly on developed clinical symptoms and typical clinical findings, e.g. the presence of autoantibodies against adrenal cortex or steroidogenic enzymes P450c17 and/or P450scc. The linkage analysis is seldom used. Further, means for natal or presymptomatic diagnosis of the disease are not easily available, since the linkage analysis provides only an indirect data through known gene markers and requires samples from several family members in several generations. Additionally, the linkage analysis is tedious and can be performed only in specialized laboratories by highly-skilled personnel.

Also the mapping of the carriers of the disease gene is presently based on the linkage analysis and thus not readily available.

Summary of the invention

We have now identified a novel gene encoding a novel zinc finger protein, designated as autoimmune regulator 1 or AIR-1, which is mutated in APECED. The novel gene and protein allow further development of the diagnosis and therapy of APECED.

The object of the invention is to provide means which are useful in a diagnostic method and a gene therapeutic method in the diagnosis and treatment of APECED.

Another object of the invention is to provide a novel method for the diagnosis APECED, including the pre- and postnatal diagnosis of and the mapping of the carriers, the method being easy and reliable to perform.

The present invention relates to an isolated DNA sequence comprising the sequence id. no. 1 or a fragment or variant thereof, or an isolated DNA sequence hybridizable thereto, the DNA sequence being associated with APECED. Preferably said isolated DNA sequence includes a gene defect responsible for APECED.

The present invention also relates to a protein comprising the amino acid sequence id. no. 2 or a fragment or variant thereof, the protein being associated with APECED. Said protein has distinct structural motifs, including the PHD finger motif (PHD), the LXXLL motif (L), proline-rich region (PRR), and cystein-rich region (CRR).

The present invention further relates to a method for the diagnosis of APECED comprising detecting in a biological specimen the presence of a DNA sequence comprising the sequence id. no. 1 or a functional fragment or variant thereof, or a DNA-sequence hybridizable thereto, the DNA sequence being associated with APECED.

The present invention further relates to the use of the above-identified DNA-sequences in the diagnosis of APECED.

The present invention further relates to a method for the diagnosis of APECED comprising detecting in a biological specimen the presence or the absence of a protein comprising the sequence id. no. 2 or a fragment thereof, the protein being associated with APECED.

The present invention further relates to the use of the above-identified protein or a fragment thereof in the diagnosis of APECED.

The present invention further relates to the use of the above-identified DNA sequences in gene therapy or for

the preparation of a pharmaceutical preparation useful in a gene therapy method of APECED.

Brief description of the drawings

Figure 1 shows a physical map of the APECED gene locus in the chromosome 21q22.3. Cosmids D1G8, D40G11, D9G11, D28B11, and D4G11, overlapping clones used for the genomic sequencing [Kudoh, J. *et al.*, DNA Res. 4 (1997) 45 -52] are indicated by horizontal lines. The APECED gene located just proximal to the 5' end of the neighboring gene PFKL is indicated by a solid arrow. N indicates *Not*I sites. DNA marker D21S1912 is shown as open box.

Figure 2 shows the structures of the APECED gene and AIR proteins. (A) Cloning strategy of AIR cDNAs and the order of the exons in the APECED gene. DNA fragments amplified by PCR and 3'- and 5'-RACE are indicated by the lines. Exon 1' is the 5'-noncoding exon of the AIR-2 and AIR-3. An additional alternative splicing of AIR-3 in exon 10, resulting in an amino acid change in its downstream, is indicated by vertical lines. Each exon, except exon 1', is bordered by the common splice site consensus sequence, ag:gt. Mutations in the exon 2 and exon 6 are indicated by the arrows. (B) Schematic presentation of the three AIR proteins showing distinct structural motifs, including the PHD finger motif (PHD), the LXXLL motif (L), proline-rich region (PRR), and cystein-rich region (CRR).

Figure 3 shows electropherograms showing the sequence surrounding the mutations in the APECED gene. (A) Mutation analysis of a Swiss APECED family. The parents are heterozygous for the allele (normal "C" and abnormal "T"). The affected boy and girl show the "C" to "T" transition resulting in the "Arg" to "Stop" nonsense mutation at amino acid position 257. (B) Mutation analysis of two Finnish APECED patients. The patient MP is homozygous for the mutant allele (left), NP is heterozygous for the allele (right). (C) The patient NP shows the "A" to "G" trans-

version resulting in the "Lys" to "Glu" missense mutation at amino acid position 42. FLEB is a normal control.

Figure 4 shows the result of restriction enzyme *Taq*I digestion assay demonstrating the R257stop mutation. Four APECED patients [HP1 (lane 1), HP2 (lane 2), NP (lane 6), and MP (lane 8)], the mothers of two families [HM (lane 5) and NM (lane 7)], two healthy siblings [HN1 (lane 3) and HN2 (lane 4)] of family H and normal controls [C1, C2 and C3 (lanes 9-11)] are shown. The APECED patients HP1, HP2 and MP are homozygotes for R257stop mutation. The APECED patient NP is heterozygous for R257stop mutation but is carrying a mutation at a different position in another allele of APECED gene (shown above in Fig. 3C). Both mothers (HM and NM) and two healthy siblings (HN1 and HN2) are heterozygous for R257stop mutation and therefore carriers of APECED but are not having the disease. Two controls (C1 and C2) are both homozygous for normal alleles. Normal alleles produce a lower 225 bp fragment, the mutated fragment is upper band at 285 bp.

Figure 5 shows an amino acid sequence alignment for the PHD finger motif of AIR-1, Mi-2, and TIF1. The consensus amino acid residues conserved in the PHD finger motif is indicated by the bold letters underneath. The residues that are identical with AIR-1 (aa 299-340) are shown by the dots. GenBank accession nos. of Mi-2 and TIF1 are X86691 and AF009353, respectively.

Figure 6. A Western blot showing the expression of AIR-1 in fetal liver. A sample of fetal liver was run on PAGE, transferred to nitrocellulose filter and probed with sera as follows: Lane 1 control mouse serum, lane 2, control mouse serum absorbed with peptide AIR-1/2 (sequence id. no. 25), lanes 3 and 4, serum from a mouse immunized with peptide AIR-1/2 for four and six weeks, respectively and absorbed with peptide AIR-1/2, lanes 5 and 6 unabsorbed serum from a mouse immunized with peptide AIR-1/2 for four

and six weeks, respectively. The strong band seen in lanes 5 and 6 represent the AIR-1 protein with a molecular weight of approx. 58 kD, the lower band is an approx. 20 kD breakdown product of the AIR protein. The bands seen in all lanes are non-specific.

Detailed description of the invention

The present invention is based on studies aiming for the identification and characterization of the gene defect in APECED. In the sequence studies, a cosmid/BAC (bacterial artificial chromosome) contig of 520 kb covering four gene markers D21S1460-D21S1912-PFKL-D21S154 [Kudoh, J. *et al.*, DNA Res. 4 (1997) 45-52] was constructed, and genomic sequencing in this region was performed [Kawasaki, K. et al., Genome Res. 7 (1997) 250-261]. From this genomic sequence information the distance between D21S1912 and PFKL was determined to be approximately 140 kb (Fig. 1).

Using a computer program, such as GRAIL and GENSCAN [Uberbacher, E. C. and Mural, R. J., Proc. Natl Acad. Sci. USA 88 (1991) 11261-11265; Burge, C. and Karlin, S., J. Mol. Biol. 268 (1997) 78-94], gene screening in the partial sequencing data within this region was performed. GENSCAN predicted several genes between D21S1912 and PFKL. One of these genes located just proximal to the PFKL gene contained the previously trapped exon HC21EXc33 [Kudoh, J. *et al.*, DNA Res. 4 (1997) 45-52] or MDC04M06 [Chen, H. *et al.*, Genome Res. 6 (1996) 747-760]. A set of primers for polymerase chain reaction (PCR) was then designed from the predicted exons. The PCR screening of various cDNA libraries using these primers allowed the isolation of a cDNA clone containing the exon HC21EXc33 (exon 13) from the thymus cDNA library (Fig. 2A).

A 3'-rapid amplification of cDNA ends (3'-RACE) and 5'-RACE using Marathon™ cDNA Amplification Kit (Clontech Laboratories Inc, California, USA) according to manufacturer's protocol from the thymus cDNA library was

performed using a primer c33F (sequence id. no. 7) and a primer 1R (sequence id. no. 8), respectively.

Sequencing analysis revealed a unique sequence of 2027 bp in overlapping PCR products that contains a 1635-bp open reading frame (ORF) from methionine at nt 128 to a TAG stop codon at nt 1763 encoding a predicted novel protein designated AIR-1, for autoimmune regulator 1. AIR-1 encodes a protein of 545 amino acids with a predicted isoelectric point of 7.32 and a calculated molecular mass of 57,723 (Fig. 2B).

A 5'-RACE from the thymus cDNA using a primer 4R (sequence id. no. 9) resulted in an alternatively spliced product. Furthermore, two types of the cDNA clones were amplified with a primer pair 3F/c33R (sequence id. no. 10/sequence id. no. 11) and these clones encode for AIR-2 and AIR-3 proteins sequence id. no. 4 and sequence id. no. 6, respectively (Fig. 2A) (sequence id. no. 3 and sequence id. no. 5). The AIR-2 and AIR-3 proteins consist of 348 and 254 amino acids, respectively (Fig. 2B). These results suggest that the APECED gene is transcribed as at least three types of mRNA by alternative splicing and/or use of an alternative 5' exon within the gene. RT-PCR analysis [Griffin, H. G. and Griffin, A. M., PCR Technology. Current Innovations, CRC Press, 1994] revealed that the AIR-1 transcript is also expressed in fetal liver (data not shown).

The APECED gene is approximately 13-kb in length and contains 15 exons, including the exon 1' specific to AIR-2 and AIR-3. It is transcribed in the direction of centromere to telomere (Figs 1, 2A). Based on this information, PCR primers were designed to amplify each exon from the genomic DNA and a mutation analysis of Swiss and Finnish APECED families was performed. Sequence comparison identified two mutations in the APECED gene of the patients (Fig. 3). The first mutation changes an Arg codon (CGA) to a stop codon (TGA) at amino acid position 257 in exon 6.

This mutation was designated as R257stop mutation. The second mutation is a missense mutation that derived from the maternal chromosome in one Finnish patient (NP): a Lys codon (AAG) changes to a Glu codon (GAG) at amino acid position 42 in exon 2. This mutation is designated as L42E mutation (Figs 2A, 3C).

The R257stop mutation destroys a *Taq*I restriction enzyme site and the K42E mutation introduces a novel *Taq*I site. Thus these two mutations can be easily demonstrated in one or both alleles by *Taq*I digestion or by digestion using another enzyme cleaving at the recognition site 5'-TCGA-3'(Fig. 4).

The AIR-1 protein has strong homology in certain domains to the major autoantigens (Mi-2) associated with the autoimmune disease dermatomyositis [Seeig, H. P. *et al.*, Arthritis Rheum. <u>38</u> (1995) 1389-1399; Ge, Q. *et al.*, J. Clin. Invest. 96 (1995) 1730-1737], Sp140, a protein from the nuclear body, an organelle involved in the pathogenesis of certain types of leukemia, and which is also the target of antibodies in the serum of patients with the autoimmune disease primary bilary cirrhosis [Bloch, D. B. *et al.*, J. Biol. Chem. <u>271</u> (1996) 29198-29204]. In addition, the homologies extend to other nuclear proteins such as TIF1 [Le Douarin, B. *et al.*, EMBO J. <u>14</u> (1995) 2020-2033], LYSP100 [Dent, A. L. *et al.*, Blood <u>88</u> (1996) 1423-1426], and putative yeast and *C. elegans* proteins. The AIR-1 protein homologies are principally in two PHD finger motifs (amino acid 299 to 340 and 434 to 475) (Fig. 5). AIR-1 also contains a proline-rich regions (amino acid 350 to 430) (Fig. 2B). The PHD finger is a cysteine-rich structure that is distinguished from the RING finger (C3HC4) and LIM domain (C2HC5) because it contains a consensus of C4HC3. [Aasland, R. et. al., Trends Biochem. Sci. <u>20</u> (1995) 56-59]. The PHD finger motif is found in a number of chromatin-associated proteins such as HRX that is involved in the t(11:17) translocation in acute leukemia [Chaplin,

T. *et al.*, Blood 86 (1995) 2073-2076]. The proline-rich region is assumed to be involved in protein-protein interaction or DNA binding. The presence of the PHD finger and proline-rich regions indicates a function for AIRs as transcription regulatory proteins. However, the AIR proteins have no apparent nuclear translocation signal, and thus other proteins containing such signal may interact with AIR to translocate it to the nucleus. In fact, the AIR proteins also have the LXXLL motif that is a signature sequence to bind to nuclear receptors [Heery, D. M. et al., Nature 387 (1997) 733-736] (Fig. 2B).

The clinical picture of APECED and the observed immunological abnormality with strong autoimmune response towards several target organs and antigens suggest that the product of the APECED gene has a central role in immune (ontogeny) maturation and in regulation of immune response towards self and nonself.

According to the diagnostic method of the invention, the presence of the defective APECED gene can be detected from a biological sample by any known detection method suitable for detecting mutations. Such methods include the method described by Saiki et al. [Proc. Natl. Acad. Sci USA 86 (1989) 6230-6234) utilizing hybridization to an allele specific oligonucleotide probe, or modifications thereof; the method described by Newton, C. R. *et al.* [Nucl. Acids Res. 17 (1989) 2503-2516] using the DNA sequences or DNA-fragments of the invention as probes; the solid phase minisequencing method described by Syvänen *et al.* [Genomics 8 (1990) 684-692] in which use is made of a biotinylated probe; or the oligonucleotide ligation method described by Landegren, U. *et al.* [Science 241 (1988) 1077-1080]. Methods include the denaturing gradient gel electrophoresis (DGGE) [Fischer, S.G. and Lerman, L.S., PNAS 80 (1983) 1579-1583] or a modification of this method, constant denaturant gel electrophoresis (CDGE) [Hoving et al., Genes Chromosomes Cancer 5 (1992) 97-103]. The mutation

separation principle of DGGE and CDGE is based on the melting behavior of the DNA double helix of a given fragment.

Since the mutations of the APECED gene involve a site sensitive to *Taq*I digestion, the mutation are preferably detected in one or both alleles by *Taq*I digestion or by digestion using another enzyme cleaving at recognition site 5′-TCGA-3′ The chemical mismatch cleavage for mutation analysis can be used [Grompe, M. et al., Proc. Natl. Acad. Sci. USA 86(15)(1989) 5888-5892].

In the diagnostic method of the invention the biological sample can be any tissue or body fluid containing cells, such as blood, e.g. umbilical cord blood, separated blood cells, such as lymphocytes, B-cells, T-cells etc., biopsy material, such as fetal liver or thymus biopsy, sperm, saliva, etc. The biological sample can be, where necessary, pretreated in a suitable manner known to those skilled in the art.

When the DNA sequence of the present invention is used therapeutically any techniques presently available for gene therapy can be employed. Accordingly, in the technique known as *ex vivo* therapy patient cells (e.g. umbilical cord blood from the fetus) with the defective gene are taken from the patient, DNA sequences encoding the normal (healthy) gene product incorporated in a carrier vector are transducted or transfected to the cells and the cells are returned to the patient. If the techniques known as *in situ* therapy is used, the DNA sequences encoding the normal gene product are first inserted to a suitable carrier vector, and the carrier is then introduced to the affected tissue, such as peripheral blood, liver or bone marrow. The carrier vector used can be a retrovirus vector, an adeno virus vector, an adeno associated virus (AAV) vector or an eucaryotic vector. The therapy can be performed intra utero or during adult life. Depending on the cells to be treated these techniques lead either to a transient cure, where

cells from affected organ are treated, or to a permanent cure, in case of the treatment of stem cells.

The present invention provides means for an easy and more rapid diagnosis of the APECED and, specifically, enables prenatal diagnosis and carrier diagnosis. Furthermore, it provides a background for therapy.

The invention is now elucidated by the following non-limiting examples.

Example 1

Localization of the APECED gene

Genomic sequencing of cosmid DNAs was performed by the shotgun method described by Kawasaki, K. *et al.*, Genome Res. 7 (1997) 250-261. Cosmids D1G8, D40G11, D9G11, D28B11, and D4G11 and gene marker D21S1912 are described by Kudoh, J. *et al.*, DNA Res. 4 (1997) 45-52].

cDNA cloning

The phage DNAs prepared from human thymus cDNA library (Clontech, HL1127a) were used as a PCR template. 20 ng of phage DNA which represents approximately $4x10^8$ phages was added to a 10 ml of reaction mixture containing 1x buffer [16mM $(NH_4)_2SO_4$, 50mM Tris-HCl, pH 9.2, 1.75 mM $MgCl_2$, 0.001% (w/v) gelatin), 0.2 mM each of dNTPs, 1M Betaine (Sigma), 0.35 U of Tap and Pwo DNA polymerase (EXpand Long Template PCR System, Boehringer Mannheim), and 0.5 mM of each of the primers, 2F and c33R, 2F and 4R, and 2F' and 2R', respectively.

The cDNA fragment was amplified by PCR using the following conditions: 94°C for 3 min., 35 cycles of 94°C for 30 sec, 60°C for 30 sec in 2F/c33R and 2F/4R or 65°C for 30 sec in 2F'/2R', and 68°C for 90 sec. 3'- and 5'-RACE were carried out by Marathon cDNA Amplification Kit (Human Thymus; Clontech). PCR reaction was performed in 10 µl volume containing 1x buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin), 0.2 mM each of dNTPs, 0.25 U of AmpliTaq Gold polymerase (Perkin-Elmer), and 0.5 mM of each of the exon-specific primers. 3'-RACE

product was amplified by PCR with the following conditions: 95°C for 9 min., 35 cycles of 94°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec.

The cDNA fragments were sequenced by the dye deoxy terminator cycle sequencing method (according to ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit protocol P/N 402078, Perkin Elmer Corporation, California) using specific primers, 2F and c33R, and AmpliTaq/FS DNA polymerase (Perkin-Elmer), and then analyzed by using an automatic DNA sequencer (Applied Biosystems 377). Primer sequences used were

1R: 5′-GTTCCCGAGTGGAAGGCGCTGC-3′ (sequence id. no. 8)
2F: 5'-GGATTCAGACCATGTCAGCTTCA-3' (sequence id. no. 12)
3F: 5′-GAGTTCAGGTACCCAGAGATGCTG-3′ (sequence id. no. 10)
c33R: 5'-CTCGCTCAGAAGGGACTCCA-3' (sequence id. no. 11)
4R: 5'-AGGGGACAGGCAGGCCAGGT-3' (sequence id. no. 9)
2F': 5'-GTGCTGTTCAAGGACTACAAC-3' (sequence id. no. 13)
2R': 5'-TGGATGAGGATCCCCTCCACG-3' (sequence id. no. 14)
AP1: 5'-CCATCCTAATACGACTCACTATAGGGC-3' (sequence id. no. 15) and
c33F: 5'-GATGACACTGCCAGTCACGA-3' (sequence id. no. 7).

Example 2

Mutation analysis of the APECED gene

For the mutation analysis the DNA samples were purified from periferal blood mononuclear cells from patients with APECED and from suspected carriers of APECED and from normal healthy controls (according to Sambrook et al. 1989, Molecular Cloning. A Laboratory Manual. CSH Press) and subjected to PCR using primers specific for all identified exons.

For sequencing the mutated exons, PCR fragments, 6F/6R in exon 6 and 49300F/49622R in exon 2, were amplified by PCR with the following conditions: 95°C for 9 min., 35 cycles of 94°C for 30 sec, 60°C for 30 sec and 72°C for 30

sec, and 94°C for 3 min., 35 cycles of 94°C for 30 sec, 60°C for 30 sec, and 68°C for 30 sec, respectively. The PCR products were sequenced using specific primers
6F: 5'-TGCAGGCTGTGGGAACTCCA-3' (sequence id. no. 16)
6R: 5'-AGAAAAAGAGCTGTACCCTGTG-3' (sequence id. no. 17)
3R: 5'-TGCAAGGAAGAGGGGCGTCAGC-3' (sequence id. no. 18)
49300F: 5'-TCCACCACAAGCCGAGGAGAT-3' (sequence id. no. 19) and 49622R: 5'-ACGGGCTCCTCAAACACCACT-3' (sequence id. no. 20).

In the mutation analysis by sequencing, two Swiss and three Finnish (HP1, HP2 and MP) patients with APECED were homozygous for R257stop allele, whereas one Finnish patient (NP) was heterozygous for this mutation (Fig. 3A, B). The R257stop mutation of NP was derived from the paternal chromosome. The second mutation, L42E mutation, was found in one Finnish patient (NP): a Lys codon (AAG) changes to a Glu codon (GAG) at amino acid position 42 in exon 2.(Figs 2A, 3C). This mutation derived from the maternal chromosome.

<u>**Example 3**</u>

Restriction enzyme TaqI analysis of two mutations in exons 2 and 6 of APECED gene

Analysis of the mutation sites in exons 2 and 6 in large series of individuals was performed using the restriction enzyme TaqI. The TaqI digestion for exons 2 and 6 was done as follows. Ten microlitres of amplification product was incubated at 65 °C for 1 hour in 20µl of reaction mixture containing 1x TaqI digestion buffer (New England Biolabs, NY, 100 µl/ml of BSA and 10u of TaqI enzyme (New England Biolabs, NY). After the digestion fragments were separated in 1,5% agarose gel and visualized by EtBr staining.

For exon 2, the fragment containing the mutation site L42E was amplified with primers GR1/2F and GR1/2R with the following conditions: 95°C for 3 min., 35 cycles of 94°C

for 30 sec, 62°C for 30 sec and 72°C for 1 min. The 1x reaction mix used contained 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin), 0.2 mM each of dNTPs, 0.25 U of Dynazyme (Finnzymes, Finland), and 0.5 mM of each of the exon-specific primers. The normal allele produces a 312 bp fragment whereas the mutated allele gives a 133 bp and a 179 bp fragment. Primer sequences for GR1/2F and GR1/2R are 5′-TGGAGATGGGCAGGCCGCAGGGTG (sequence id. no. 21) and 5′-CAGTCCAGCTGGGCTGAGCAGGTG (sequence id. no. 22), respectively.

For exon 6, the fragment containing the R257stop mutation site was amplified with primers GR1/5IF and GR1/5IR with the same conditions described for exon 2 (see above). The normal allele produces a 225 bp fragment whereas the mutated allele gives a 285 bp fragment. Primer sequences for GR1/5IF and GR1/5IR are 5'GCGGCTCCAAGAAGTG-CATCCAGG (sequence id. no. 23) and 5'-CTCCACCCTGCAAGGAA-GAGGGGC (sequence id. no. 24), respectively.

The screening of 50 Finnish and 50 Swiss healthy individuals did not reveal R257stop or K42E mutations by *Taq*I digestion. Similarly, PCR analysis of 20 unaffected Japanese was performed and no mutations were found in any of these positions. These results demonstrate that the APECED gene is responsible for the pathogenesis of APECED.

Mutations were found in the AIR-1 transcript but not in the AIR-2 and AIR-3 transcripts from all the APECED patients tested. Two Swiss and three Finnish (HP1, HP2 and MP) patients who are homozygous for the R257stop mutation completely lack functional AIR-1 protein but still have intact AIR-2 and AIR-3 proteins.

One common mutation seems responsible for the genetic defect in approximately 90% of the Finnish APECED cases and a haplotype analysis with the markers D21S141, D21S1912 and PFKL shows that the R257stop mutation is

likely to be this common mutation [Björses, P. *et al.*, Am. J. Hum. Genet. 59 (1996) 879-886].

Example 4

Analysis of the AIR protein expression

In this example, synthetic peptides representing amino-acid sequences of the AIR-1 protein, were used to generate a polyvalent mouse antiserum against the AIR-1 protein.

For the peptide synthesis, two peptides were chosen according to the antigenicity prediction by Pepsort program (GCC package, Wisconsin, USA). The peptides AIR-1/2 and AIR-1/6 (TLHLKEKEGCPQAFH, sequence id. no. 25 and GKNKARSSSGPKPLV, sequence id. no. 26, respectively) representing exons 2 and 6, respectively, of the APECED gene were synthesized onto a branched lysine core (Fmoc8-Lys4-Lys2-Lys-betaAla-Wang resin, Calbiochem-Novabiochem, La Jolla, Ca, USA) resulting in an octameric multible antigen peptide (MAP) [Tam, J. P. *et al.*, Proc. Natl. Acad. Sci. USA 85 (1988) 5409-5413; Adermann, K. *et al.*, in Solid Phase Synthesis, Biological and Biomedical Applications, pp. 429-432, Ed. R. Epton, Mayflower Worldwide Ltd., Birmingham, 1994], Syntheses were performed by Fmoc (N-(9-fluorenyl)methoxycarbonyl) chemistry on a simultaneous multiple peptide synthesizer (SMPS 350, Zinsser Analytic, Frankfurt, Germany). Purity of MAPs was analyzed by reverse-phase HPLC (System Gold, Beckman Instruments Inc, Fullerton, CA, USA).

To obtain murine polyclonal antibodies, eight-week old Balb/c mice were immunized with an intraperitoneal injection of 25 micrograms of each peptide in 0,4 ml of a 1:1 mixture of Freund's Complete Adjuvant (Difco Laboratories, Detroit, MI, USA) and physiological saline (NaCl, 0,15 M). One month later the animals were boosted with an intramuscular injection of 35 micrograms of antigens in Freund's incomplete adjuvant and saline (1:1) (0,2 ml were distributed into four sites). Three weeks

later the peptides in a dose of 50 micrograms/mouse were administered intravenously and sera were obtained 7 days later.

For the production of EBV transformed B-cells, peripheral blood leukocytes were obtained from healthy control persons. The B-cells were transformed with EBV (Epstein-Barr virus) using standard protocol, and the cell lines were maintained in RPMI 1640, supplemented with 10% FCS (fetal calf serum). An aliquot of cells were stimulated for 12 hours with 10 μg/ml of phytohemagglutinin (PHA) to obtain mitogen-activated T-cells.

Tissue samples were obtained from stillborn fetuses at six months gestational age. Fetal liver, spleen, thymus and lymphnodes were homogenized, the homogenates were cleared with centrifugations (20 000 rpm for 20 minutes) and the samples were used for western blot analysis.

For analysis of polyclonal sera, Elisa and western blot analysis were performed. Microtitre ELISA plates (Maxisorp, Nunc, Roskilde, Denmark) were coated with the peptides (1 micrograms /well in PBS, pH 7,5) at 4°C overnight and blocked with 2 % of BSA in PBS. The plates were then incubated with titrated mouse immune sera and normal (control) sera at room temperature for 4 h. Finally the bound peptide-specific antibodies were detected by use of anti-mouse HRP-labelled immunoglobulins (Dako A/S, Denmark) essentially as previously described [Ovod, V. A. *et al.*, AIDS 6 (1992) 25.34].

For western blotting, tissue homogenates, EBV transformed B-cells or PHA-activated T-cells were boiled for 10 minutes in 2x sample buffer (for tissue homogenates: 100 microliters of homogenate mixed with 100 microliters of sample buffer. For cells: one million cells/100 μl of buffer) and analyzed in western blotting as described in Ovod, V. A. *et al., supra*.

The antisera so produced reacted with the AIR-1- protein low amount in normal fetal spleen, thymus and

lymphonode as well as, in EBV-transformed B-cells and in PHA-activated T-cells. In the ELISA assay towards the immunogenic peptides, all four mice gave a strong reactivity towards the peptide used for the immunization. In the western blotting analysis using either the tissue homogenates or stimulated T-cells or established B-cells, a strong band of approx. 60 kD molecular weight was seen in fetal liver (Fig. 6), while weaker bands of the same size were seen in the other samples.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Kai Krohn et al.
(B) STREET: Iltarusko, Salmentaantie 751
(C) CITY: 36450 Salmentaka
(E) COUNTRY: Finland
(F) POSTAL CODE (ZIP): none

(ii) TITLE OF INVENTION: Novel Gene

(iii) NUMBER OF SEQUENCES: 26

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2036 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:137..1774
(D) OTHER INFORMATION:/product= "AIR-1"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:137..1771
(D) OTHER INFORMATION:/product= "AIR-1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
AGACCGGGGA GACGGGCGGG CGCACAGCCG GCGCGGAGGC CCCACAGCCC CGCCGGGACC      60

CGAGGCCAAG CGAGGGGCTG CCAGTGTCCC GGGACCCACC GCGTCCGCCC CAGCCCCGGG     120

TCCCCGCGCC CACCCC ATG GCG ACG GAC GCG GCG CTA CGC CGG CTT CTG         169
                  Met Ala Thr Asp Ala Ala Leu Arg Arg Leu Leu
                   1               5                  10

AGG CTG CAC CGC ACG GAG ATC GCG GTG GCC GTG GAC AGC GCC TTC CCA       217
Arg Leu His Arg Thr Glu Ile Ala Val Ala Val Asp Ser Ala Phe Pro
            15                  20                  25

CTG CTG CAC GCG CTG GCT GAC CAC GAC GTG GTC CCC GAG GAC AAG TTT       265
Leu Leu His Ala Leu Ala Asp His Asp Val Val Pro Glu Asp Lys Phe
        30                  35                  40
```

```
CAG GAG ACG CTT CAT CTG AAG GAA AAG GAG GGC TGC CCC CAG GCC TTC      313
Gln Glu Thr Leu His Leu Lys Glu Lys Glu Gly Cys Pro Gln Ala Phe
     45                 50                  55

CAC GCC CTC CTG TCC TGG CTG CTG ACC CAG GAC TCC ACA GCC ATC CTG      361
His Ala Leu Leu Ser Trp Leu Leu Thr Gln Asp Ser Thr Ala Ile Leu
 60                  65                  70                  75

GAC TTC TGG AGG GTG CTG TTC AAG GAC TAC AAC CTG GAG CGC TAT GGC      409
Asp Phe Trp Arg Val Leu Phe Lys Asp Tyr Asn Leu Glu Arg Tyr Gly
                 80                  85                  90

CGG CTG CAG CCC ATC CTG GAC AGC TTC CCC AAA GAT GTG GAC CTC AGC      457
Arg Leu Gln Pro Ile Leu Asp Ser Phe Pro Lys Asp Val Asp Leu Ser
             95                 100                 105

CAG CCC CGG AAG GGG AGG AAG CCC CCG GCC GTC CCC AAG GCT TTG GTA      505
Gln Pro Arg Lys Gly Arg Lys Pro Pro Ala Val Pro Lys Ala Leu Val
        110                 115                 120

CCG CCA CCC AGA CTC CCC ACC AAG AGG AAG GCC TCA GAA GAG GCT CGA      553
Pro Pro Pro Arg Leu Pro Thr Lys Arg Lys Ala Ser Glu Glu Ala Arg
    125                 130                 135

GCT GCC GCG CCA GCA GCC CTG ACT CCA AGG GGC ACC GCC AGC CCA GGC      601
Ala Ala Ala Pro Ala Ala Leu Thr Pro Arg Gly Thr Ala Ser Pro Gly
140                 145                 150                 155

TCT CAA CTG AAG GCC AAG CCC CCC AAG AAG CCG GAG AGC AGC GCA GAG      649
Ser Gln Leu Lys Ala Lys Pro Pro Lys Lys Pro Glu Ser Ser Ala Glu
                160                 165                 170

CAG CAG CGC CTT CCA CTC GGG AAC GGG ATT CAG ACC ATG TCA GCT TCA      697
Gln Gln Arg Leu Pro Leu Gly Asn Gly Ile Gln Thr Met Ser Ala Ser
            175                 180                 185

GTC CAG AGA GCT GTG GCC ATG TCC TCC GGG GAC GTC CCG GGA GCC CGA      745
Val Gln Arg Ala Val Ala Met Ser Ser Gly Asp Val Pro Gly Ala Arg
        190                 195                 200

GGG GCC GTG GAG GGG ATC CTC ATC CAG CAG GTG TTT GAG TCA GGC GGC      793
Gly Ala Val Glu Gly Ile Leu Ile Gln Gln Val Phe Glu Ser Gly Gly
    205                 210                 215

TCC AAG AAG TGC ATC CAG GTT GGC GGG GAG TTC TAC ACT CCC AGC AAG      841
Ser Lys Lys Cys Ile Gln Val Gly Gly Glu Phe Tyr Thr Pro Ser Lys
220                 225                 230                 235

TTC GAA GAC TCC GGC AGT GGG AAG AAC AAG GCC CGC AGC AGC AGT GGC      889
Phe Glu Asp Ser Gly Ser Gly Lys Asn Lys Ala Arg Ser Ser Ser Gly
                240                 245                 250

CCG AAG CCT CTG GTT CGA GCC AAG GGA GCC CAG GGC GCT GCC CCC GGT      937
Pro Lys Pro Leu Val Arg Ala Lys Gly Ala Gln Gly Ala Ala Pro Gly
            255                 260                 265

GGA GGT GAG GCT AGG CTG GGC CAG CAG GGC AGC GTT CCC GCC CCT CTG      985
Gly Gly Glu Ala Arg Leu Gly Gln Gln Gly Ser Val Pro Ala Pro Leu
        270                 275                 280

GCC CTC CCC AGT GAC CCC CAG CTC CAC CAG AAG AAT GAG GAC GAG TGT     1033
Ala Leu Pro Ser Asp Pro Gln Leu His Gln Lys Asn Glu Asp Glu Cys
    285                 290                 295
```

GCC GTG TGT CGG GAC GGC GGG GAG CTC ATC TGC TGT GAC GGC TGC CCT 1081
Ala Val Cys Arg Asp Gly Gly Glu Leu Ile Cys Cys Asp Gly Cys Pro
300 305 310 315

CGG GCC TTC CAC CTG GCC TGC CTG TCC CCT CCG CTC CGG GAG ATC CCC 1129
Arg Ala Phe His Leu Ala Cys Leu Ser Pro Pro Leu Arg Glu Ile Pro
320 325 330

AGT GGG ACC TGG AGG TGC TCC AGC TGC CTG CAG GCA ACA GTC CAG GAG 1177
Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu Gln Ala Thr Val Gln Glu
335 340 345

GTG CAG CCC CGG GCA GAG GAG CCC CGG CCC CAG GAG CCA CCC GTG GAG 1225
Val Gln Pro Arg Ala Glu Glu Pro Arg Pro Gln Glu Pro Pro Val Glu
350 355 360

ACC CCG CTC CCC CCG GGG CTT AGG TCG GCG GGA GAG GAG GTA AGA GGT 1273
Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala Gly Glu Glu Val Arg Gly
365 370 375

CCA CCT GGG GAA CCC CTA GCC GGC ATG GAC ACG ACT CTT GTC TAC AAG 1321
Pro Pro Gly Glu Pro Leu Ala Gly Met Asp Thr Thr Leu Val Tyr Lys
380 385 390 395

CAC CTG CCG GCT CCG CCT TCT GCA GCC CCG CTG CCA GGG CTG GAC TCC 1369
His Leu Pro Ala Pro Pro Ser Ala Ala Pro Leu Pro Gly Leu Asp Ser
400 405 410

TCG GCC CTG CAC CCC CTA CTG TGT GTG GGT CCT GAG GGT CAG CAG AAC 1417
Ser Ala Leu His Pro Leu Leu Cys Val Gly Pro Glu Gly Gln Gln Asn
415 420 425

CTG GCT CCT GGT GCG CGT TGC GGG GTG TGC GGA GAT GGT ACG GAC GTG 1465
Leu Ala Pro Gly Ala Arg Cys Gly Val Cys Gly Asp Gly Thr Asp Val
430 435 440

CTG CGG TGT ACT CAC TGC GCC GCT GCC TTC CAC TGG CGC TGC CAC TTC 1513
Leu Arg Cys Thr His Cys Ala Ala Ala Phe His Trp Arg Cys His Phe
445 450 455

CCA GCC GGC ACC TCC CGG CCC GGG ACG GGC CTG CGC TGC AGA TCC TGC 1561
Pro Ala Gly Thr Ser Arg Pro Gly Thr Gly Leu Arg Cys Arg Ser Cys
460 465 470 475

TCA GGA GAC GTG ACC CCA GCC CCT GTG GAG GGG GTG CTG GCC CCC AGC 1609
Ser Gly Asp Val Thr Pro Ala Pro Val Glu Gly Val Leu Ala Pro Ser
480 485 490

CCC GCC CGC CTG GCC CCT GGG CCT GCC AAG GAT GAC ACT GCC AGT CAC 1657
Pro Ala Arg Leu Ala Pro Gly Pro Ala Lys Asp Asp Thr Ala Ser His
495 500 505

GAG CCC GCT CTG CAC AGG GAT GAC CTG GAG TCC CTT CTG AGC GAG CAC 1705
Glu Pro Ala Leu His Arg Asp Asp Leu Glu Ser Leu Leu Ser Glu His
510 515 520

ACC TTC GAT GGC ATC CTG CAG TGG GCC ATC CAG AGC ATG GCC CGT CCG 1753
Thr Phe Asp Gly Ile Leu Gln Trp Ala Ile Gln Ser Met Ala Arg Pro
525 530 535

GCG GCC CCC TTC CCC TCC TGA CCCCAGATGG CCGGGACATG CAGCTCTGAT 1804
Ala Ala Pro Phe Pro Ser *
540 545

GAGAGAGTGC TGAGAAGGAC ACCTCCTTCC TCAGTCCTGG AAGCCGGCCG GCTGGGATCA 1864

AGAAGGGGAC AGCGCCACCT CTTGTCAGTG CTCGGCTGTA AACAGCTCTG TGTTTCTGGG 1924

GACACCAGCC ATCATGTGCC TGGAAATTAA ACCCTGCCCC ACTTCTCTAC TCTGGAAGTC 1984

CCCGGGAGCC TCTCCTTGCC TGGTGACCTA CTAAAAATAT AAAAATTAGC TG 2036

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 545 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Thr Asp Ala Ala Leu Arg Arg Leu Leu Arg Leu His Arg Thr
1 5 10 15

Glu Ile Ala Val Ala Val Asp Ser Ala Phe Pro Leu Leu His Ala Leu
20 25 30

Ala Asp His Asp Val Val Pro Glu Asp Lys Phe Gln Glu Thr Leu His
35 40 45

Leu Lys Glu Lys Glu Gly Cys Pro Gln Ala Phe His Ala Leu Leu Ser
50 55 60

Trp Leu Leu Thr Gln Asp Ser Thr Ala Ile Leu Asp Phe Trp Arg Val
65 70 75 80

Leu Phe Lys Asp Tyr Asn Leu Glu Arg Tyr Gly Arg Leu Gln Pro Ile
85 90 95

Leu Asp Ser Phe Pro Lys Asp Val Asp Leu Ser Gln Pro Arg Lys Gly
100 105 110

Arg Lys Pro Pro Ala Val Pro Lys Ala Leu Val Pro Pro Pro Arg Leu
115 120 125

Pro Thr Lys Arg Lys Ala Ser Glu Glu Ala Arg Ala Ala Ala Pro Ala
130 135 140

Ala Leu Thr Pro Arg Gly Thr Ala Ser Pro Gly Ser Gln Leu Lys Ala
145 150 155 160

Lys Pro Pro Lys Lys Pro Glu Ser Ser Ala Glu Gln Gln Arg Leu Pro
165 170 175

Leu Gly Asn Gly Ile Gln Thr Met Ser Ala Ser Val Gln Arg Ala Val
180 185 190

Ala Met Ser Ser Gly Asp Val Pro Gly Ala Arg Gly Ala Val Glu Gly
195 200 205

Ile Leu Ile Gln Gln Val Phe Glu Ser Gly Gly Ser Lys Lys Cys Ile
210 215 220

Gln Val Gly Gly Glu Phe Tyr Thr Pro Ser Lys Phe Glu Asp Ser Gly
225 230 235 240

Ser Gly Lys Asn Lys Ala Arg Ser Ser Ser Gly Pro Lys Pro Leu Val
245 250 255

Arg Ala Lys Gly Ala Gln Gly Ala Ala Pro Gly Gly Gly Glu Ala Arg
260 265 270

Leu Gly Gln Gln Gly Ser Val Pro Ala Pro Leu Ala Leu Pro Ser Asp
275 280 285

Pro Gln Leu His Gln Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp
290 295 300

Gly Gly Glu Leu Ile Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu
305 310 315 320

Ala Cys Leu Ser Pro Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg
325 330 335

Cys Ser Ser Cys Leu Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala
340 345 350

Glu Glu Pro Arg Pro Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro
355 360 365

Gly Leu Arg Ser Ala Gly Glu Glu Val Arg Gly Pro Pro Gly Glu Pro
370 375 380

Leu Ala Gly Met Asp Thr Thr Leu Val Tyr Lys His Leu Pro Ala Pro
385 390 395 400

Pro Ser Ala Ala Pro Leu Pro Gly Leu Asp Ser Ser Ala Leu His Pro
405 410 415

Leu Leu Cys Val Gly Pro Glu Gly Gln Gln Asn Leu Ala Pro Gly Ala
420 425 430

Arg Cys Gly Val Cys Gly Asp Gly Thr Asp Val Leu Arg Cys Thr His
435 440 445

Cys Ala Ala Ala Phe His Trp Arg Cys His Phe Pro Ala Gly Thr Ser
450 455 460

Arg Pro Gly Thr Gly Leu Arg Cys Arg Ser Cys Ser Gly Asp Val Thr
465 470 475 480

Pro Ala Pro Val Glu Gly Val Leu Ala Pro Ser Pro Ala Arg Leu Ala
485 490 495

Pro Gly Pro Ala Lys Asp Asp Thr Ala Ser His Glu Pro Ala Leu His
500 505 510

Arg Asp Asp Leu Glu Ser Leu Leu Ser Glu His Thr Phe Asp Gly Ile
515 520 525

Leu Gln Trp Ala Ile Gln Ser Met Ala Arg Pro Ala Ala Pro Phe Pro
530 535 540

Ser *
545

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1545 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:237..1283
(D) OTHER INFORMATION:/product= "AIR-2"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:237..1280
(D) OTHER INFORMATION:/product= "AIR-2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AGAGAAAGTG AGGTCTTCTC AGGCTCTTAA GAGCATGGCG TTTGGTCCAG GCTGTACCCG 60

CTGCTCTCAG CTGGGCCCGT GGGTGGGCCG GGCGCCCCTG CTATAGCCAG GAGGTCAAGG 120

ATCCACTGGG AATGCCATGC TCATCTTTCG TCCCCAGCAT GGTTTCTTAA TGGGGTAGAA 180

GCAGGTCGGG AGAGACCTCC CTGGGCCTGG CCCCACTGCC CTGTGAGGAA GGGTTC 236

ATG TGG TTG GTG TAC AGT TCC GGG GCC CCT GGA ACG CAG CAG CCT GCA 284
Met Trp Leu Val Tyr Ser Ser Gly Ala Pro Gly Thr Gln Gln Pro Ala
1 5 10 15

AGA AAC CGG GTT TTC TTC CCA ATA GGG ATG GCC CCG GGG GGT GTC TGT 332
Arg Asn Arg Val Phe Phe Pro Ile Gly Met Ala Pro Gly Gly Val Cys
20 25 30

TGG AGA CCA GAT GGA TGG GGA ACA GGT GGT CAG GGC AGA ATT TCA GGC 380
Trp Arg Pro Asp Gly Trp Gly Thr Gly Gly Gln Gly Arg Ile Ser Gly
35 40 45

CCT GGC AGC ATG GGA GCA GGG CAG AGA CTG GGG AGT TCA GGT ACC CAG 428
Pro Gly Ser Met Gly Ala Gly Gln Arg Leu Gly Ser Ser Gly Thr Gln
50 55 60

AGA TGC TGC TGG GGG AGC TGT TTT GGG AAG GAG GTG GCT CTC AGG AGG 476
Arg Cys Cys Trp Gly Ser Cys Phe Gly Lys Glu Val Ala Leu Arg Arg
65 70 75 80

GTG CTG CAC CCC AGC CCA GTC TGC ATG GGC GTC TCT TGC CTG TGC CAG 524
Val Leu His Pro Ser Pro Val Cys Met Gly Val Ser Cys Leu Cys Gln
85 90 95

AAG AAT GAG GAC GAG TGT GCC GTG TGT CGG GAC GGC GGG GAG CTC ATC 572
Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp Gly Gly Glu Leu Ile
100 105 110

TGC TGT GAC GGC TGC CCT CGG GCC TTC CAC CTG GCC TGC CTG TCC CCT 620
Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu Ala Cys Leu Ser Pro
115 120 125

```
CCG CTC CGG GAG ATC CCC AGT GGG ACC TGG AGG TGC TCC AGC TGC CTG        668
Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu
    130                 135                 140

CAG GCA ACA GTC CAG GAG GTG CAG CCC CGG GCA GAG GAG CCC CGG CCC        716
Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala Glu Glu Pro Arg Pro
145                 150                 155                 160

CAG GAG CCA CCC GTG GAG ACC CCG CTC CCC CCG GGG CTT AGG TCG GCG        764
Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala
                165                 170                 175

GGA GAG GAG GTA AGA GGT CCA CCT GGG GAA CCC CTA GCC GGC ATG GAC        812
Gly Glu Glu Val Arg Gly Pro Pro Gly Glu Pro Leu Ala Gly Met Asp
            180                 185                 190

ACG ACT CTT GTC TAC AAG CAC CTG CCG GCT CCG CCT TCT GCA GCC CCG        860
Thr Thr Leu Val Tyr Lys His Leu Pro Ala Pro Pro Ser Ala Ala Pro
        195                 200                 205

CTG CCA GGG CTG GAC TCC TCG GCC CTG CAC CCC CTA CTG TGT GTG GGT        908
Leu Pro Gly Leu Asp Ser Ser Ala Leu His Pro Leu Leu Cys Val Gly
    210                 215                 220

CCT GAG GGT CAG CAG AAC CTG GCT CCT GGT GCG CGT TGC GGG GTG TGC        956
Pro Glu Gly Gln Gln Asn Leu Ala Pro Gly Ala Arg Cys Gly Val Cys
225                 230                 235                 240

GGA GAT GGT ACG GAC GTG CTG CGG TGT ACT CAC TGC GCC GCT GCC TTC       1004
Gly Asp Gly Thr Asp Val Leu Arg Cys Thr His Cys Ala Ala Ala Phe
                245                 250                 255

CAC TGG CGC TGC CAC TTC CCA GCC GGC ACC TCC CGG CCC GGG ACG GGC       1052
His Trp Arg Cys His Phe Pro Ala Gly Thr Ser Arg Pro Gly Thr Gly
            260                 265                 270

CTG CGC TGC AGA TCC TGC TCA GGA GAC GTG ACC CCA GCC CCT GTG GAG       1100
Leu Arg Cys Arg Ser Cys Ser Gly Asp Val Thr Pro Ala Pro Val Glu
        275                 280                 285

GGG GTG CTG GCC CCC AGC CCC GCC CGC CTG GCC CCT GGG CCT GCC AAG       1148
Gly Val Leu Ala Pro Ser Pro Ala Arg Leu Ala Pro Gly Pro Ala Lys
    290                 295                 300

GAT GAC ACT GCC AGT CAC GAG CCC GCT CTG CAC AGG GAT GAC CTG GAG       1196
Asp Asp Thr Ala Ser His Glu Pro Ala Leu His Arg Asp Asp Leu Glu
305                 310                 315                 320

TCC CTT CTG AGC GAG CAC ACC TTC GAT GGC ATC CTG CAG TGG GCC ATC       1244
Ser Leu Leu Ser Glu His Thr Phe Asp Gly Ile Leu Gln Trp Ala Ile
                325                 330                 335

CAG AGC ATG GCC CGT CCG GCG GCC CCC TTC CCC TCC TGA CCCCAGATGG        1293
Gln Ser Met Ala Arg Pro Ala Ala Pro Phe Pro Ser  *
            340                 345

CCGGGACATG CAGCTCTGAT GAGAGAGTGC TGAGAAGGAC ACCTCCTTCC TCAGTCCTGG     1353

AAGCCGGCCG GCTGGGATCA AGAAGGGGAC AGCGCCACCT CTTGTCAGTG CTCGGCTGTA     1413

AACAGCTCTG TGTTTCTGGG GACACCAGCC ATCATGTGCC TGGAAATTAA ACCCTGCCCC     1473
```

ACTTCTCTAC TCTGGAAGTC CCCGGGAGCC TCTCCTTGCC TGGTGACCTA CTAAAAATAT 1533

AAAAATTAGC TG 1545

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 348 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Trp Leu Val Tyr Ser Ser Gly Ala Pro Gly Thr Gln Gln Pro Ala
  1               5                  10                  15

Arg Asn Arg Val Phe Phe Pro Ile Gly Met Ala Pro Gly Gly Val Cys
             20                  25                  30

Trp Arg Pro Asp Gly Trp Gly Thr Gly Gly Gln Gly Arg Ile Ser Gly
         35                  40                  45

Pro Gly Ser Met Gly Ala Gly Gln Arg Leu Gly Ser Ser Gly Thr Gln
     50                  55                  60

Arg Cys Cys Trp Gly Ser Cys Phe Gly Lys Glu Val Ala Leu Arg Arg
 65                  70                  75                  80

Val Leu His Pro Ser Pro Val Cys Met Gly Val Ser Cys Leu Cys Gln
                 85                  90                  95

Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp Gly Gly Glu Leu Ile
            100                 105                 110

Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu Ala Cys Leu Ser Pro
        115                 120                 125

Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu
    130                 135                 140

Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala Glu Glu Pro Arg Pro
145                 150                 155                 160

Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala
                165                 170                 175

Gly Glu Glu Val Arg Gly Pro Pro Gly Glu Pro Leu Ala Gly Met Asp
            180                 185                 190

Thr Thr Leu Val Tyr Lys His Leu Pro Ala Pro Pro Ser Ala Ala Pro
        195                 200                 205

Leu Pro Gly Leu Asp Ser Ser Ala Leu His Pro Leu Leu Cys Val Gly
    210                 215                 220

Pro Glu Gly Gln Gln Asn Leu Ala Pro Gly Ala Arg Cys Gly Val Cys
225                 230                 235                 240

Gly Asp Gly Thr Asp Val Leu Arg Cys Thr His Cys Ala Ala Ala Phe
                245                 250                 255
```

```
His Trp Arg Cys His Phe Pro Ala Gly Thr Ser Arg Pro Gly Thr Gly
            260                 265                 270

Leu Arg Cys Arg Ser Cys Ser Gly Asp Val Thr Pro Ala Pro Val Glu
        275                 280                 285

Gly Val Leu Ala Pro Ser Pro Ala Arg Leu Ala Pro Gly Pro Ala Lys
    290                 295                 300

Asp Asp Thr Ala Ser His Glu Pro Ala Leu His Arg Asp Asp Leu Glu
305                 310                 315                 320

Ser Leu Leu Ser Glu His Thr Phe Asp Gly Ile Leu Gln Trp Ala Ile
                325                 330                 335

Gln Ser Met Ala Arg Pro Ala Ala Pro Phe Pro Ser  *
            340                 345
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1463 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION:237..1001
- (D) OTHER INFORMATION:/product= "AIR-3"

(ix) FEATURE:
- (A) NAME/KEY: mat_peptide
- (B) LOCATION:237..998
- (D) OTHER INFORMATION:/product= "AIR-3"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
AGAGAAAGTG AGGTCTTCTC AGGCTCTTAA GAGCATGGCG TTTGGTCCAG GCTGTACCCG     60

CTGCTCTCAG CTGGGCCCGT GGGTGGGCCG GGCGCCCCTG CTATAGCCAG GAGGTCAAGG    120

ATCCACTGGG AATGCCATGC TCATCTTTCG TCCCCAGCAT GGTTTCTTAA TGGGGTAGAA    180

GCAGGTCGGG AGAGACCTCC CTGGGCCTGG CCCCACTGCC CTGTGAGGAA GGGTTC        236

ATG TGG TTG GTG TAC AGT TCC GGG GCC CCT GGA ACG CAG CAG CCT GCA      284
Met Trp Leu Val Tyr Ser Ser Gly Ala Pro Gly Thr Gln Gln Pro Ala
  1               5                  10                  15

AGA AAC CGG GTT TTC TTC CCA ATA GGG ATG GCC CCG GGG GGT GTC TGT      332
Arg Asn Arg Val Phe Phe Pro Ile Gly Met Ala Pro Gly Gly Val Cys
             20                  25                  30

TGG AGA CCA GAT GGA TGG GGA ACA GGT GGT CAG GGC AGA ATT TCA GGC      380
Trp Arg Pro Asp Gly Trp Gly Thr Gly Gly Gln Gly Arg Ile Ser Gly
         35                  40                  45

CCT GGC AGC ATG GGA GCA GGG CAG AGA CTG GGG AGT TCA GGT ACC CAG      428
Pro Gly Ser Met Gly Ala Gly Gln Arg Leu Gly Ser Ser Gly Thr Gln
     50                  55                  60
```

AGA TGC TGC TGG GGG AGC TGT TTT GGG AAG GAG GTG GCT CTC AGG AGG 476
Arg Cys Cys Trp Gly Ser Cys Phe Gly Lys Glu Val Ala Leu Arg Arg
65 70 75 80

GTG CTG CAC CCC AGC CCA GTC TGC ATG GGC GTC TCT TGC CTG TGC CAG 524
Val Leu His Pro Ser Pro Val Cys Met Gly Val Ser Cys Leu Cys Gln
85 90 95

AAG AAT GAG GAC GAG TGT GCC GTG TGT CGG GAC GGC GGG GAG CTC ATC 572
Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp Gly Gly Glu Leu Ile
100 105 110

TGC TGT GAC GGC TGC CCT CGG GCC TTC CAC CTG GCC TGC CTG TCC CCT 620
Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu Ala Cys Leu Ser Pro
115 120 125

CCG CTC CGG GAG ATC CCC AGT GGG ACC TGG AGG TGC TCC AGC TGC CTG 668
Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu
130 135 140

CAG GCA ACA GTC CAG GAG GTG CAG CCC CGG GCA GAG GAG CCC CGG CCC 716
Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala Glu Glu Pro Arg Pro
145 150 155 160

CAG GAG CCA CCC GTG GAG ACC CCG CTC CCC CCG GGG CTT AGG TCG GCG 764
Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala
165 170 175

GGA GAG GAG CCC CGC TGC CAG GGC TGG ACT CCT CGG CCC TGC ACC CCC 812
Gly Glu Glu Pro Arg Cys Gln Gly Trp Thr Pro Arg Pro Cys Thr Pro
180 185 190

TAC TGT GTG TGG GTC CTG AGG GTC AGC AGA ACC TGG CTC CTG GTG CGC 860
Tyr Cys Val Trp Val Leu Arg Val Ser Arg Thr Trp Leu Leu Val Arg
195 200 205

GTT GCG GGG TGT GCG GAG ATG GTA CGG ACG TGC TGC GGT GTA CTC ACT 908
Val Ala Gly Cys Ala Glu Met Val Arg Thr Cys Cys Gly Val Leu Thr
210 215 220

GCG CCG CTG CCT TCC ACT GGC GCT GCC ACT TCC CAG CCG GCA CCT CCC 956
Ala Pro Leu Pro Ser Thr Gly Ala Ala Thr Ser Gln Pro Ala Pro Pro
225 230 235 240

GGC CCG GGA CGG GCC TGC GCT GCA GAT CCT GCT CAG GAG ACG TGA 1001
Gly Pro Gly Arg Ala Cys Ala Ala Asp Pro Ala Gln Glu Thr *
245 250 255

CCCCAGCCCC TGTGGAGGGG GTGCTGGCCC CCAGCCCCGC CCGCCTGGCC CCTGGGCCTG 1061

CCAAGGATGA CACTGCCAGT CACGAGCCCG CTCTGCACAG GGATGACCTG GAGTCCCTTC 1121

TGAGCGAGCA CACCTTCGAT GGCATCCTGC AGTGGGCCAT CCAGAGCATG GCCCGTCCGG 1181

CGGCCCCCTT CCCCTCCTGA CCCCAGATGG CCGGGACATG CAGCTCTGAT GAGAGAGTGC 1241

TGAGAAGGAC ACCTCCTTCC TCAGTCCTGG AAGCCGGCCG GCTGGGATCA AGAAGGGGAC 1301

AGCGCCACCT CTTGTCAGTG CTCGGCTGTA AACAGCTCTG TGTTTCTGGG GACACCAGCC 1361

ATCATGTGCC TGGAAATTAA ACCCTGCCCC ACTTCTCTAC TCTGGAAGTC CCCGGGAGCC 1421

TCTCCTTGCC TGGTGACCTA CTAAAAATAT AAAAATTAGC TG 1463

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 254 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Trp Leu Val Tyr Ser Ser Gly Ala Pro Gly Thr Gln Gln Pro Ala
 1               5                  10                  15

Arg Asn Arg Val Phe Phe Pro Ile Gly Met Ala Pro Gly Gly Val Cys
            20                  25                  30

Trp Arg Pro Asp Gly Trp Gly Thr Gly Gly Gln Gly Arg Ile Ser Gly
        35                  40                  45

Pro Gly Ser Met Gly Ala Gly Gln Arg Leu Gly Ser Ser Gly Thr Gln
    50                  55                  60

Arg Cys Cys Trp Gly Ser Cys Phe Gly Lys Glu Val Ala Leu Arg Arg
65                  70                  75                  80

Val Leu His Pro Ser Pro Val Cys Met Gly Val Ser Cys Leu Cys Gln
                85                  90                  95

Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp Gly Gly Glu Leu Ile
            100                 105                 110

Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu Ala Cys Leu Ser Pro
        115                 120                 125

Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu
    130                 135                 140

Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala Glu Glu Pro Arg Pro
145                 150                 155                 160

Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala
                165                 170                 175

Gly Glu Glu Pro Arg Cys Gln Gly Trp Thr Pro Arg Pro Cys Thr Pro
            180                 185                 190

Tyr Cys Val Trp Val Leu Arg Val Ser Arg Thr Trp Leu Leu Val Arg
        195                 200                 205

Val Ala Gly Cys Ala Glu Met Val Arg Thr Cys Cys Gly Val Leu Thr
    210                 215                 220

Ala Pro Leu Pro Ser Thr Gly Ala Ala Thr Ser Gln Pro Ala Pro Pro
225                 230                 235                 240

Gly Pro Gly Arg Ala Cys Ala Ala Asp Pro Ala Gln Glu Thr  *
                245                 250                 255
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GATGACACTG CCAGTCACGA 20

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTTCCCGAGT GGAAGGCGCT GC 22

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AGGGGACAGG CAGGCCAGGT 20

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GAGTTCAGGT ACCCAGAGAT GCTG 24

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CTCGCTCAGA AGGGACTCCA 20

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGATTCAGAC CATGTCAGCT TCA 23

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GTGCTGTTCA AGGACTACAA C 21

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TGGATGAGGA TCCCCTCCAC G 21

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCATCCTAAT ACGACTCACT ATAGGGC 27

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TGCAGGCTGT GGGAACTCCA 20

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AGAAAAAGAG CTGTACCCTG TG 22

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

TGCAAGGAAG AGGGGCGTCA GC 22

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TCCACCACAA GCCGAGGAGA T 21

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ACGGGCTCCT CAAACACCAC T 21

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TGGAGATGGG CAGGCCGCAG GGTG 24

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CAGTCCAGCT GGGCTGAGCA GGTG 24

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GCGGCTCCAA GAAGTGCATC CAGG 24

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CTCCACCCTG CAAGGAAGAG GGGC 24

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Thr Leu His Leu Lys Glu Lys Glu Gly Cys Pro Gln Ala Phe His
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Gly Lys Asn Lys Ala Arg Ser Ser Ser Gly Pro Lys Pro Leu Val
1 5 10 15

Claims

1. An isolated DNA sequence characterized by comprising the sequence id. no. 1 or a fragment or variant thereof, or an isolated DNA sequence hybridizable thereto, the DNA sequence being associated with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED).

2. An isolated DNA sequence according to claim 1, characterized in that it includes a gene defect responsible for APECED.

3. A DNA sequence according to claim 1, characterized by having the sequence according to sequence id. no 1 or a fragment thereof having the sequence according to sequence id. no 3 or sequence id. no 5.

4. A protein characterized by comprising the amino acid sequence id. no. 2 or a fragment or variant thereof, the protein being associated with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED).

5. A protein according to claim 4 characterized by having the amino acid sequence id. no. 2, or a fragment thereof having the sequence according to sequence id. no. 4, or a fragment thereof having the sequence according to sequence id. no 6.

6. A protein according to claim 4 or 5 characterized by having distinct structural motifs, including the PHD finger motif (PHD), the LXXLL motif (L), proline-rich region (PRR), and cystein-rich region (CRR).

7. A method for the diagnosis of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) characterized by detecting in a biological specimen the precense of a DNA sequence comprising the sequence id. no. 1 or a functional fragment or variant thereof, or an isolated DNA-sequence hybridizable thereto, the DNA sequence being associated with APECED.

8. A method according to claim 7, characterized in that the DNA sequence includes a gene defect responsible for APECED.

9. A method according to claim 8, characterized in that the gene defect to be detected includes a "C" to "T" transition resulting in the "Arg" to "Stop" nonsense mutation at amino acid position 257 and/or a "A" to "G" transversion resulting in the "Lys" to "Glu" missense mutation at amino acid position 42.

10. A method according to any one of claims 7 to 9, characterized in that DNA techniques are used for the detection.

11. A method according to any one of claims 7 to 10, characterized in that the detection takes advantage of TaqI or another enzyme cleaving at recognition site 5'-TCGA-3' digestion.

12. A method for the diagnosis of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) characterized by detecting in a biological specimen the precense or the absence of a protein comprising the sequence id. no. 1, or a fragment thereof having the sequence according to sequence id. no. 4, or a fragment therof having the sequence according to sequence id. no 6, the protein being associated with APECED.

13. The use of the DNA sequence according to any one of claims 1 to 3 in the diagnosis of APECED.

14. The use of the protein according to any one of claims 4 to 6 in the diagnosis of APECED.

15. The use of the DNA sequence according to any one of claims 1 to 3 for the preparation of a medicament useful in a gene therapy method of APECED.

16. The use of the DNA sequence according to any one of claims 1 to 3 in the treatment of APECED.

(57) Abstract

The present invention relates to a novel gene, a novel protein encoded by said gene, a mutated form of the gene and to diagnostic and therapeutic uses of the gene or a mutated form thereof. More specifically, the present invention relates to a novel gene defective in autoimmune polyendocrinopathy syndrome type I (APS I), also called autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) (MIM No. 240,300).

ん3

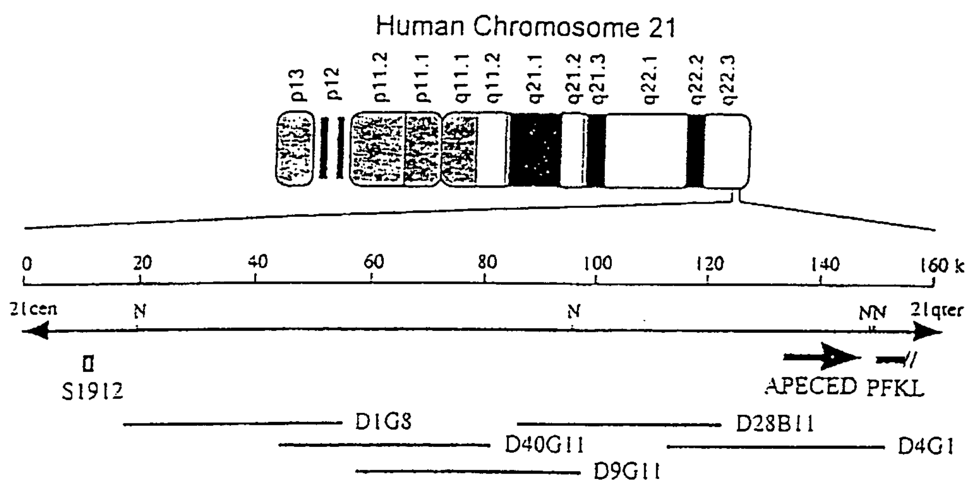
Human Chromosome 21
p13
p12
p11.2
p11.1
q11.1
q11.2
q21.1
q21.2
q21.3
q22.1
q22.2
q22.3
0
20
40
60
80
100
120
140
160 k
21cen
N
N
NN
21qter
S1912
APECED
PFKL
D1G8
D28B11
D40G11
D4G1
D9G11

Fig. 1

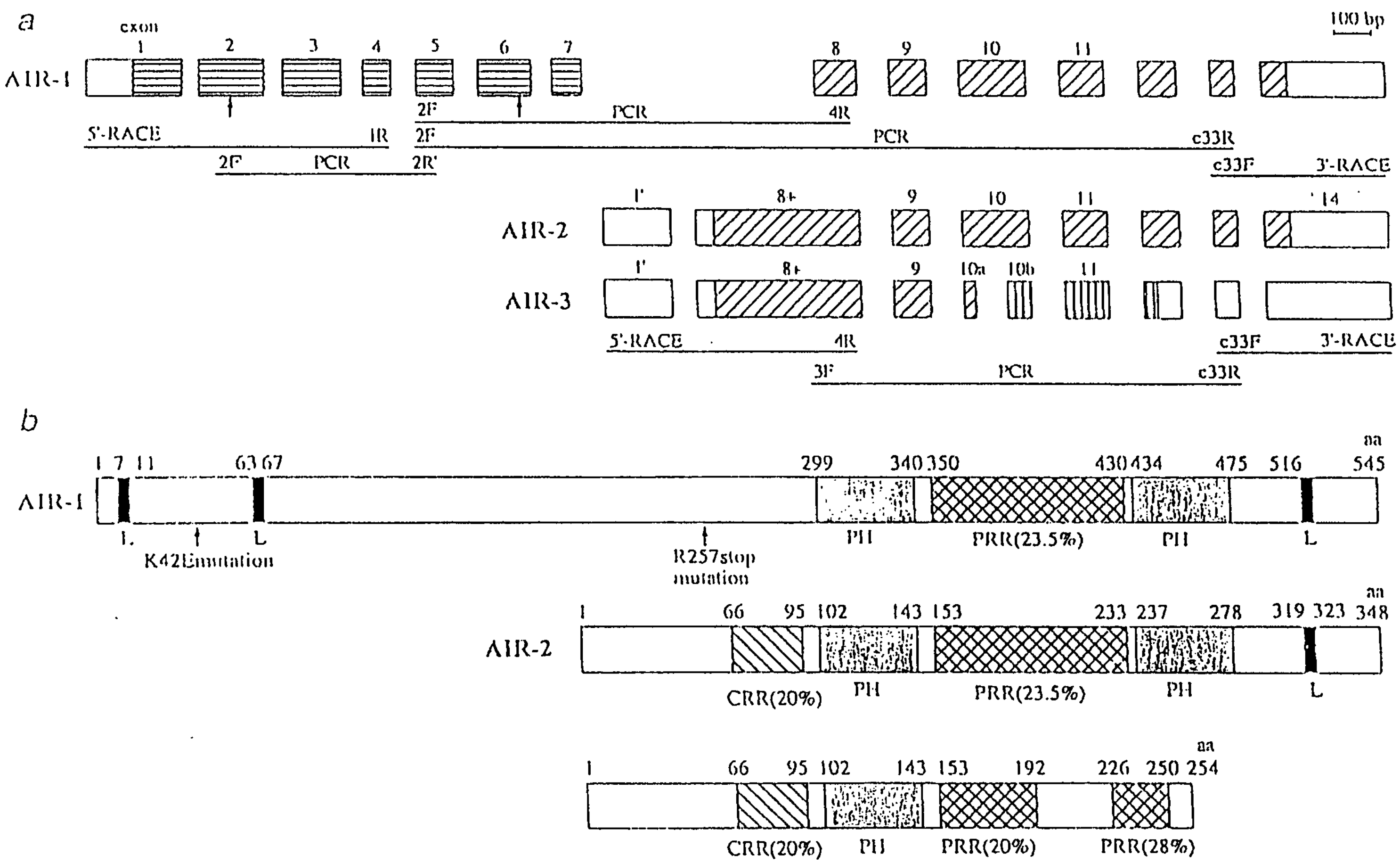
a
exon
AIR-1
1
2
3
4
5
6
7
8
9
10
11
100 bp
5'-RACE
2F
PCR
1R
2R'
2F
PCR
4R
2F
PCR
c33R
c33F
3'-RACE
AIR-2
1'
8+
9
10
11
14
AIR-3
1'
8+
9
10a
10b
11
5'-RACE
4R
3F
PCR
c33R
c33F
3'-RACE
b
AIR-1
1 7 11
63 67
299
340 350
430 434
475
516
aa
545
L
K42Emutation
L
R257stop mutation
PH
PRR(23.5%)
PH
L
AIR-2
1
66
95 102
143 153
233 237
278
319 323
aa
348
CRR(20%)
PH
PRR(23.5%)
PH
L
1
66
95 102
143 153
192
226
250 254
aa
CRR(20%)
PH
PRR(20%)
PRR(28%)

Fig. 2

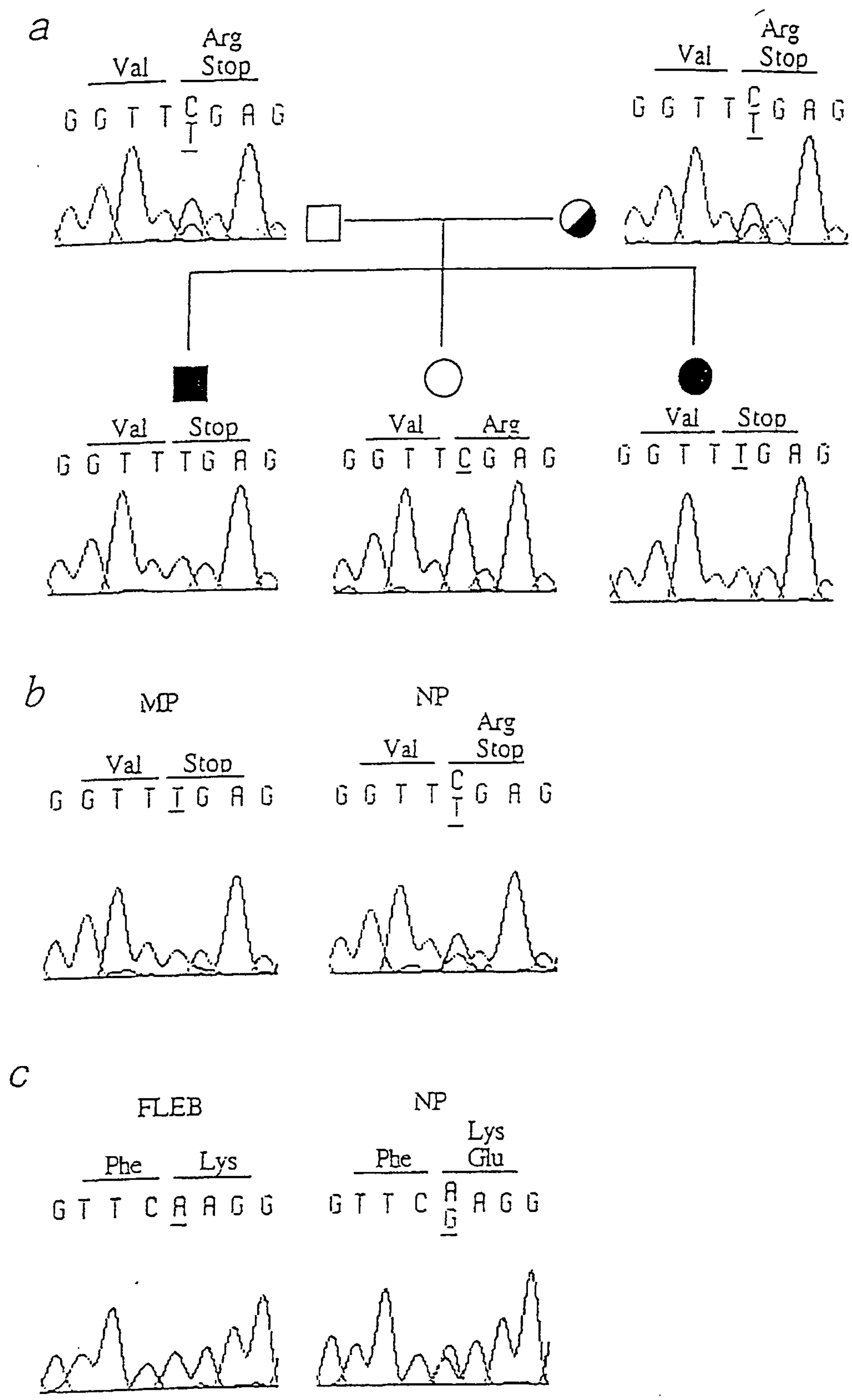
a
Val
Arg
Stop
G G T T C/T G A G
Val
Stop
G G T T T G A G
Val
Arg
G G T T C G A G
Val
Stop
G G T T T G A G
b
MP
Val
Stop
G G T T T G A G
NP
Val
Arg
Stop
G G T T C/T G A G
c
FLEB
Phe
Lys
G T T C A A G G
NP
Phe
Lys
Glu
G T T C A/G A G G

Fig. 3

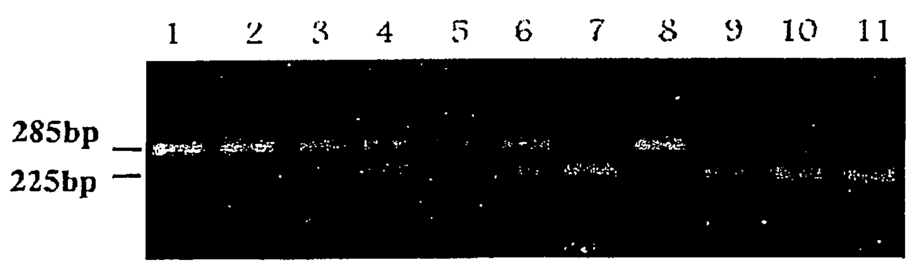

Fig.4

```
AIR-1: 299 CAVCRDGGELICCDGCPRAFHLACLSPPLREIPSGTWRCSSC 340
AIR-1: 434 .G..G..TDVLR.TH.AA...WR.HF.AGTSR.GTGL..R.. 475
Mi-2 : 373 .E..QQ...I.L..T....Y.MV..D.DMEKA.E.K.S.PH. 414
Mi-2 : 452 .R..K.........T..SSY.IH..N...P...N.E.L.PR. 493
TIF1 : 791 ....QN.......EK..KV...S.HV.T.TNF...E.I.TF. 832
consensus  C  C        C  C    H  C              C  C
```

Fig. 5

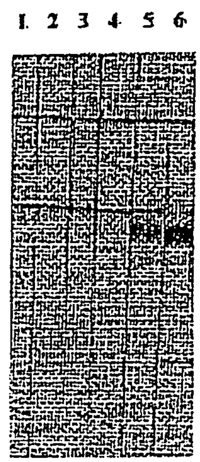

Fig. 6

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

OIPE IAP53
OCT 02 2006
PATENT & TRADEMARK OFFICE

In re application of: KAI KROHN, ET AL.

Serial No.: 09/508,658 Group No.: 1634

Filed: NOVEMBER 3, 2000 Examiner: SITTON, J.S.

For: NOVEL GENE DEFECTIVE IN APECED AND ITS USE

Attorney Docket No.: U 012653-9

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

ATTACHMENT 1

Attached are SEQ ID Nos: 1 of U.S. Patent Application 09/508,658 and U.S. Patent 6,951,928 that show that the nucleic acid that can be a "T" rather than a "C" is found at the same location in the nucleotide sequence encoding SEQ ID NO:2. This is the nucleotide 768 of the coding sequence.

Respectfully submitted,

JANET I. CORD
LADAS & PARRY LLP
26 WEST 61 STREET
NEW YORK, NEW YORK 10023
REG. NO: 33778 (212) 708-1935

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Kai Krohn et al.
(B) STREET: Iltarusko, Salmentaantie 751
(C) CITY: 36450 Salmentaka
(E) COUNTRY: Finland
(F) POSTAL CODE (ZIP): none

(ii) TITLE OF INVENTION: Novel Gene

(iii) NUMBER OF SEQUENCES: 26

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2036 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:137..1774
(D) OTHER INFORMATION:/product= "AIR-1"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:137..1771
(D) OTHER INFORMATION:/product= "AIR-1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
AGACCGGGGA GACGGGCGGG CGCACAGCCG GCGCGGAGGC CCCACAGCCC CGCCGGGACC      60

CGAGGCCAAG CGAGGGGCTG CCAGTGTCCC GGGACCCACC GCGTCCGCCC CAGCCCCGGG     120

TCCCCGCGCC CACCCC ATG GCG ACG GAC GCG GCG CTA CGC CGG CTT CTG         169
                  Met Ala Thr Asp Ala Ala Leu Arg Arg Leu Leu
                   1               5                  10
```

AGG CTG CAC CGC ACG GAG ATC GCG GTG GCC GTG GAC AGC GCC TTC CCA 217
Arg Leu His Arg Thr Glu Ile Ala Val Ala Val Asp Ser Ala Phe Pro
15 20 25

CTG CTG CAC GCG CTG GCT GAC CAC GAC GTG GTC CCC GAG GAC AAG TTT 265
Leu Leu His Ala Leu Ala Asp His Asp Val Val Pro Glu Asp Lys Phe
30 35 40

CAG GAG ACG CTT CAT CTG AAG GAA AAG GAG GGC TGC CCC CAG GCC TTC 313
Gln Glu Thr Leu His Leu Lys Glu Lys Glu Gly Cys Pro Gln Ala Phe
45 50 55

CAC GCC CTC CTG TCC TGG CTG CTG ACC CAG GAC TCC ACA GCC ATC CTG 361
His Ala Leu Leu Ser Trp Leu Leu Thr Gln Asp Ser Thr Ala Ile Leu
60 65 70 75

GAC TTC TGG AGG GTG CTG TTC AAG GAC TAC AAC CTG GAG CGC TAT GGC 409
Asp Phe Trp Arg Val Leu Phe Lys Asp Tyr Asn Leu Glu Arg Tyr Gly
80 85 90

CGG CTG CAG CCC ATC CTG GAC AGC TTC CCC AAA GAT GTG GAC CTC AGC 457
Arg Leu Gln Pro Ile Leu Asp Ser Phe Pro Lys Asp Val Asp Leu Ser
95 100 105

CAG CCC CGG AAG GGG AGG AAG CCC CCG GCC GTC CCC AAG GCT TTG GTA 505
Gln Pro Arg Lys Gly Arg Lys Pro Pro Ala Val Pro Lys Ala Leu Val
110 115 120

CCG CCA CCC AGA CTC CCC ACC AAG AGG AAG GCC TCA GAA GAG GCT CGA 553
Pro Pro Pro Arg Leu Pro Thr Lys Arg Lys Ala Ser Glu Glu Ala Arg
125 130 135

GCT GCC GCG CCA GCA GCC CTG ACT CCA AGG GGC ACC GCC AGC CCA GGC 601
Ala Ala Ala Pro Ala Ala Leu Thr Pro Arg Gly Thr Ala Ser Pro Gly
140 145 150 155

TCT CAA CTG AAG GCC AAG CCC CCC AAG AAG CCG GAG AGC AGC GCA GAG 649
Ser Gln Leu Lys Ala Lys Pro Pro Lys Lys Pro Glu Ser Ser Ala Glu
160 165 170

CAG CAG CGC CTT CCA CTC GGG AAC GGG ATT CAG ACC ATG TCA GCT TCA 697
Gln Gln Arg Leu Pro Leu Gly Asn Gly Ile Gln Thr Met Ser Ala Ser
175 180 185

GTC CAG AGA GCT GTG GCC ATG TCC TCC GGG GAC GTC CCG GGA GCC CGA 745
Val Gln Arg Ala Val Ala Met Ser Ser Gly Asp Val Pro Gly Ala Arg
190 195 200

GGG GCC GTG GAG GGG ATC CTC ATC CAG CAG GTG TTT GAG TCA GGC GGC 793
Gly Ala Val Glu Gly Ile Leu Ile Gln Gln Val Phe Glu Ser Gly Gly
205 210 215

TCC AAG AAG TGC ATC CAG GTT GGC GGG GAG TTC TAC ACT CCC AGC AAG 841
Ser Lys Lys Cys Ile Gln Val Gly Gly Glu Phe Tyr Thr Pro Ser Lys
220 225 230 235

TTC GAA GAC TCC GGC AGT GGG AAG AAC AAG GCC CGC AGC AGC AGT GGC 889
Phe Glu Asp Ser Gly Ser Gly Lys Asn Lys Ala Arg Ser Ser Ser Gly
240 245 250

CCG AAG CCT CTG GTT CGA GCC AAG GGA GCC CAG GGC GCT GCC CCC GGT 937
Pro Lys Pro Leu Val Arg Ala Lys Gly Ala Gln Gly Ala Ala Pro Gly
255 260 265

GGA GGT GAG GCT AGG CTG GGC CAG CAG GGC AGC GTT CCC GCC CCT CTG 985
Gly Gly Glu Ala Arg Leu Gly Gln Gln Gly Ser Val Pro Ala Pro Leu
270 275 280

GCC CTC CCC AGT GAC CCC CAG CTC CAC CAG AAG AAT GAG GAC GAG TGT 1033
Ala Leu Pro Ser Asp Pro Gln Leu His Gln Lys Asn Glu Asp Glu Cys
285 290 295

GCC GTG TGT CGG GAC GGC GGG GAG CTC ATC TGC TGT GAC GGC TGC CCT 1081
Ala Val Cys Arg Asp Gly Gly Glu Leu Ile Cys Cys Asp Gly Cys Pro
300 305 310 315

CGG GCC TTC CAC CTG GCC TGC CTG TCC CCT CCG CTC CGG GAG ATC CCC 1129
Arg Ala Phe His Leu Ala Cys Leu Ser Pro Pro Leu Arg Glu Ile Pro
320 325 330

AGT GGG ACC TGG AGG TGC TCC AGC TGC CTG CAG GCA ACA GTC CAG GAG 1177
Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu Gln Ala Thr Val Gln Glu
335 340 345

GTG CAG CCC CGG GCA GAG GAG CCC CGG CCC CAG GAG CCA CCC GTG GAG 1225
Val Gln Pro Arg Ala Glu Glu Pro Arg Pro Gln Glu Pro Pro Val Glu
350 355 360

ACC CCG CTC CCC CCG GGG CTT AGG TCG GCG GGA GAG GAG GTA AGA GGT 1273
Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala Gly Glu Glu Val Arg Gly
365 370 375

CCA CCT GGG GAA CCC CTA GCC GGC ATG GAC ACG ACT CTT GTC TAC AAG 1321
Pro Pro Gly Glu Pro Leu Ala Gly Met Asp Thr Thr Leu Val Tyr Lys
380 385 390 395

CAC CTG CCG GCT CCG CCT TCT GCA GCC CCG CTG CCA GGG CTG GAC TCC 1369
His Leu Pro Ala Pro Pro Ser Ala Ala Pro Leu Pro Gly Leu Asp Ser
400 405 410

TCG GCC CTG CAC CCC CTA CTG TGT GTG GGT CCT GAG GGT CAG CAG AAC 1417
Ser Ala Leu His Pro Leu Leu Cys Val Gly Pro Glu Gly Gln Gln Asn
415 420 425

CTG GCT CCT GGT GCG CGT TGC GGG GTG TGC GGA GAT GGT ACG GAC GTG 1465
Leu Ala Pro Gly Ala Arg Cys Gly Val Cys Gly Asp Gly Thr Asp Val
430 435 440

CTG CGG TGT ACT CAC TGC GCC GCT GCC TTC CAC TGG CGC TGC CAC TTC 1513
Leu Arg Cys Thr His Cys Ala Ala Ala Phe His Trp Arg Cys His Phe
445 450 455

CCA GCC GGC ACC TCC CGG CCC GGG ACG GGC CTG CGC TGC AGA TCC TGC 1561
Pro Ala Gly Thr Ser Arg Pro Gly Thr Gly Leu Arg Cys Arg Ser Cys
460 465 470 475

TCA GGA GAC GTG ACC CCA GCC CCT GTG GAG GGG GTG CTG GCC CCC AGC 1609
Ser Gly Asp Val Thr Pro Ala Pro Val Glu Gly Val Leu Ala Pro Ser
480 485 490

CCC GCC CGC CTG GCC CCT GGG CCT GCC AAG GAT GAC ACT GCC AGT CAC 1657
Pro Ala Arg Leu Ala Pro Gly Pro Ala Lys Asp Asp Thr Ala Ser His
495 500 505

GAG CCC GCT CTG CAC AGG GAT GAC CTG GAG TCC CTT CTG AGC GAG CAC 1705
Glu Pro Ala Leu His Arg Asp Asp Leu Glu Ser Leu Leu Ser Glu His
510 515 520

ACC TTC GAT GGC ATC CTG CAG TGG GCC ATC CAG AGC ATG GCC CGT CCG 1753
Thr Phe Asp Gly Ile Leu Gln Trp Ala Ile Gln Ser Met Ala Arg Pro
525 530 535

GCG GCC CCC TTC CCC TCC TGA CCCCAGATGG CCGGGACATG CAGCTCTGAT 1804
Ala Ala Pro Phe Pro Ser *
540 545

GAGAGAGTGC TGAGAAGGAC ACCTCCTTCC TCAGTCCTGG AAGCCGGCCG GCTGGGATCA 1864

AGAAGGGGAC AGCGCCACCT CTTGTCAGTG CTCGGCTGTA AACAGCTCTG TGTTTCTGGG 1924

GACACCAGCC ATCATGTGCC TGGAAATTAA ACCCTGCCCC ACTTCTCTAC TCTGGAAGTC 1984

CCCGGGAGCC TCTCCTTGCC TGGTGACCTA CTAAAAATAT AAAAATTAGC TG 2036

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 545 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Thr Asp Ala Ala Leu Arg Arg Leu Leu Arg Leu His Arg Thr
1 5 10 15

Glu Ile Ala Val Ala Val Asp Ser Ala Phe Pro Leu Leu His Ala Leu
20 25 30

Ala Asp His Asp Val Val Pro Glu Asp Lys Phe Gln Glu Thr Leu His
35 40 45

SEQUENCE LISTING

```
<160> NUMBER OF SEQ ID NOS: 30

<210> SEQ ID NO 1
<211> LENGTH: 2245
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (121)..(1758)
<223> OTHER INFORMATION:

<400> SEQUENCE: 1

cgggcgcaca gccggcgcgg aggccccaca gccccgccgg gacccgaggc caagcgaggg      60

gctgccagtg tcccgggacc caccgcgtcc gccccagccc cgggtccccg cgcccacccc     120

atg gcg acg gac gcg gcg cta cgc cgg ctt ctg agg ctg cac cgc acg       168
Met Ala Thr Asp Ala Ala Leu Arg Arg Leu Leu Arg Leu His Arg Thr
1               5                   10                  15

gag atc gcg gtg gcc gtg gac agc gcc ttc cca ctg ctg cac gcg ctg       216
Glu Ile Ala Val Ala Val Asp Ser Ala Phe Pro Leu Leu His Ala Leu
            20                  25                  30

gct gac cac gac gtg gtc ccc gag gac aag ttt cag gag acg ctt cat       264
Ala Asp His Asp Val Val Pro Glu Asp Lys Phe Gln Glu Thr Leu His
        35                  40                  45

ctg aag gaa aag gag ggc tgc ccc cag gcc ttc cac gcc ctc ctg tcc       312
Leu Lys Glu Lys Glu Gly Cys Pro Gln Ala Phe His Ala Leu Leu Ser
    50                  55                  60

tgg ctg ctg acc cag gac tcc aca gcc atc ctg gac ttc tgg agg gtg       360
Trp Leu Leu Thr Gln Asp Ser Thr Ala Ile Leu Asp Phe Trp Arg Val
65                  70                  75                  80

ctg ttc aag gac tac aac ctg gag cgc tat ggc cgg ctg cag ccc atc       408
Leu Phe Lys Asp Tyr Asn Leu Glu Arg Tyr Gly Arg Leu Gln Pro Ile
                85                  90                  95

ctg gac agc ttc ccc aaa gat gtg gac ctc agc cag ccc cgg aag ggg       456
Leu Asp Ser Phe Pro Lys Asp Val Asp Leu Ser Gln Pro Arg Lys Gly
            100                 105                 110

agg aag ccc ccg gcc gtc ccc aag gct ttg gta ccg cca ccc aga ctc       504
Arg Lys Pro Pro Ala Val Pro Lys Ala Leu Val Pro Pro Pro Arg Leu
        115                 120                 125

ccc acc aag agg aag gcc tca gaa gag gct cga gct gcc gcg cca gca       552
Pro Thr Lys Arg Lys Ala Ser Glu Glu Ala Arg Ala Ala Ala Pro Ala
    130                 135                 140

gcc ctg act cca agg ggc acc gcc agc cca ggc tct caa ctg aag gcc       600
Ala Leu Thr Pro Arg Gly Thr Ala Ser Pro Gly Ser Gln Leu Lys Ala
145                 150                 155                 160

aag ccc ccc aag aag ccg gag agc agc gca gag cag cag cgc ctt cca       648
Lys Pro Pro Lys Lys Pro Glu Ser Ser Ala Glu Gln Gln Arg Leu Pro
                165                 170                 175

ctc ggg aac ggg att cag acc atg tca gct tca gtc cag aga gct gtg       696
Leu Gly Asn Gly Ile Gln Thr Met Ser Ala Ser Val Gln Arg Ala Val
            180                 185                 190

gcc atg tcc tcc ggg gac gtc ccg gga gcc cga ggg gcc gtg gag ggg       744
Ala Met Ser Ser Gly Asp Val Pro Gly Ala Arg Gly Ala Val Glu Gly
        195                 200                 205

atc ctc atc cag cag gtg ttt gag tca ggc ggc tcc aag aag tgc atc       792
Ile Leu Ile Gln Gln Val Phe Glu Ser Gly Gly Ser Lys Lys Cys Ile
    210                 215                 220

cag gtt ggt ggg gag ttc tac act ccc agc aag ttc gaa gac tcc ggc       840
Gln Val Gly Gly Glu Phe Tyr Thr Pro Ser Lys Phe Glu Asp Ser Gly
225                 230                 235                 240
```

-continued

```
agt ggg aag aac aag gcc cgc agc agc agt ggc ccg aag cct ctg gtt      888
Ser Gly Lys Asn Lys Ala Arg Ser Ser Ser Gly Pro Lys Pro Leu Val
            245                 250                 255

cga gcc aag gga gcc cag ggc gct gcc ccc ggt gga ggt gag gct agg      936
Arg Ala Lys Gly Ala Gln Gly Ala Ala Pro Gly Gly Gly Glu Ala Arg
        260                 265                 270

ctg ggc cag cag ggc agc gtt ccc gcc cct ctg gcc ctc ccc agt gac      984
Leu Gly Gln Gln Gly Ser Val Pro Ala Pro Leu Ala Leu Pro Ser Asp
    275                 280                 285

ccc cag ctc cac cag aag aat gag gac gag tgt gcc gtg tgt cgg gac     1032
Pro Gln Leu His Gln Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp
290                 295                 300

ggc ggg gag ctc atc tgc tgt gac ggc tgc cct cgg gcc ttc cac ctg     1080
Gly Gly Glu Leu Ile Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu
305                 310                 315                 320

gcc tgc ctg tcc cct ccg ctc cgg gag atc ccc agt ggg acc tgg agg     1128
Ala Cys Leu Ser Pro Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg
            325                 330                 335

tgc tcc agc tgc ctg cag gca aca gtc cag gag gtg cag ccc cgg gca     1176
Cys Ser Ser Cys Leu Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala
        340                 345                 350

gag gag ccc cgg ccc cag gag cca ccc gtg gag acc ccg ctc ccc ccg     1224
Glu Glu Pro Arg Pro Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro
    355                 360                 365

ggg ctt agg tcg gcg gga gag gag gta aga ggt cca cct ggg gaa ccc     1272
Gly Leu Arg Ser Ala Gly Glu Glu Val Arg Gly Pro Pro Gly Glu Pro
    370                 375                 380

cta gcc ggc atg gac acg act ctt gtc tac aag cac ctg ccg gct ccg     1320
Leu Ala Gly Met Asp Thr Thr Leu Val Tyr Lys His Leu Pro Ala Pro
385                 390                 395                 400

cct tct gca gcc ccg ctg cca ggg ctg gac tcc tcg gcc ctg cac ccc     1368
Pro Ser Ala Ala Pro Leu Pro Gly Leu Asp Ser Ser Ala Leu His Pro
            405                 410                 415

cta ctg tgt gtg ggt cct gag ggt cag cag aac ctg gct cct ggt gcg     1416
Leu Leu Cys Val Gly Pro Glu Gly Gln Gln Asn Leu Ala Pro Gly Ala
        420                 425                 430

cgt tgc ggg gtg tgc gga gat ggt acg gac gtg ctg cgg tgt act cac     1464
Arg Cys Gly Val Cys Gly Asp Gly Thr Asp Val Leu Arg Cys Thr His
    435                 440                 445

tgc gcc gct gcc ttc cac tgg cgc tgc cac ttc cca gcc ggc acc tcc     1512
Cys Ala Ala Ala Phe His Trp Arg Cys His Phe Pro Ala Gly Thr Ser
    450                 455                 460

cgg ccc ggg acg ggc ctg cgc tgc aga tcc tgc tca gga gac gtg acc     1560
Arg Pro Gly Thr Gly Leu Arg Cys Arg Ser Cys Ser Gly Asp Val Thr
465                 470                 475                 480

cca gcc cct gtg gag ggg gtg ctg gcc ccc agc ccc gcc cgc ctg gcc     1608
Pro Ala Pro Val Glu Gly Val Leu Ala Pro Ser Pro Ala Arg Leu Ala
            485                 490                 495

cct ggg cct gcc aag gat gac act gcc agt cac gag ccc gct ctg cac     1656
Pro Gly Pro Ala Lys Asp Asp Thr Ala Ser His Glu Pro Ala Leu His
        500                 505                 510

agg gat gac ctg gag tcc ctt ctg agc gag cac acc ttc gat ggc atc     1704
Arg Asp Asp Leu Glu Ser Leu Leu Ser Glu His Thr Phe Asp Gly Ile
    515                 520                 525

ctg cag tgg gcc atc cag agc atg gcc cgt ccg gcg gcc ccc ttc ccc     1752
Leu Gln Trp Ala Ile Gln Ser Met Ala Arg Pro Ala Ala Pro Phe Pro
    530                 535                 540

tcc tga ccccagatgg ccgggacatg cagctctgat gagagagtgc tgagaaggac      1808
Ser
```

-continued

545

acctccttcc tcagtcctgg aagccggccg gctgggatca agaaggggac agcgccacct 1868

cttgtcagtg ctcggctgta aacagctctg tgtttctggg gacaccagcc atcatgtgcc 1928

tggaaattaa accctgcccc acttctctac tctggaagtc cccgggagcc tctccttgcc 1988

tggtgaccta ctaaaaatat aaaaattagc tgggtgtggt ggtgggtgcc tgtaatccca 2048

gctacatggg agcctgaggc atgagaatca cttgaactcg ggaggtggag gttgcagtga 2108

gctgagattg cgccactgca ctccagtctg gtcggcaaga gtgagactcc gtctcaaaaa 2168

caaaacaaaa aaaccacata acataaattt atcatctcga ccacttttca gttcagtggc 2228

attcacatct catgtaa 2245

<210> SEQ ID NO 2
<211> LENGTH: 545
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

Met Ala Thr Asp Ala Ala Leu Arg Arg Leu Leu Arg Leu His Arg Thr
1 5 10 15

Glu Ile Ala Val Ala Val Asp Ser Ala Phe Pro Leu Leu His Ala Leu
20 25 30

Ala Asp His Asp Val Val Pro Glu Asp Lys Phe Gln Glu Thr Leu His
35 40 45

Leu Lys Glu Lys Glu Gly Cys Pro Gln Ala Phe His Ala Leu Leu Ser
50 55 60

Trp Leu Leu Thr Gln Asp Ser Thr Ala Ile Leu Asp Phe Trp Arg Val
65 70 75 80

Leu Phe Lys Asp Tyr Asn Leu Glu Arg Tyr Gly Arg Leu Gln Pro Ile
85 90 95

Leu Asp Ser Phe Pro Lys Asp Val Asp Leu Ser Gln Pro Arg Lys Gly
100 105 110

Arg Lys Pro Pro Ala Val Pro Lys Ala Leu Val Pro Pro Pro Arg Leu
115 120 125

Pro Thr Lys Arg Lys Ala Ser Glu Glu Ala Arg Ala Ala Ala Pro Ala
130 135 140

Ala Leu Thr Pro Arg Gly Thr Ala Ser Pro Gly Ser Gln Leu Lys Ala
145 150 155 160

Lys Pro Pro Lys Lys Pro Glu Ser Ser Ala Glu Gln Gln Arg Leu Pro
165 170 175

Leu Gly Asn Gly Ile Gln Thr Met Ser Ala Ser Val Gln Arg Ala Val
180 185 190

Ala Met Ser Ser Gly Asp Val Pro Gly Ala Arg Gly Ala Val Glu Gly
195 200 205

Ile Leu Ile Gln Gln Val Phe Glu Ser Gly Gly Ser Lys Lys Cys Ile
210 215 220

Gln Val Gly Gly Glu Phe Tyr Thr Pro Ser Lys Phe Glu Asp Ser Gly
225 230 235 240

Ser Gly Lys Asn Lys Ala Arg Ser Ser Ser Gly Pro Lys Pro Leu Val
245 250 255

Arg Ala Lys Gly Ala Gln Gly Ala Ala Pro Gly Gly Gly Glu Ala Arg
260 265 270

Leu Gly Gln Gln Gly Ser Val Pro Ala Pro Leu Ala Leu Pro Ser Asp
275 280 285